

# A SYSTEMATIC RETRIEVAL ANALYSIS OF SECONDARY ECLIPSE SPECTRA I: A COMPARISON OF ATMOSPHERIC RETRIEVAL TECHNIQUES

MICHAEL R. LINE

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125

AARON S. WOLF

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125

XI ZHANG

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125

HEATHER KNUTSON

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125

JOSHUA A. KAMMER

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125

ELIAS ELLISON

Flintridge Preparatory School, La Cañada, CA 91011

PIETER DEROO

Jet Propulsion Laboratory, California Institute of Technology

DAVE CRISP

Jet Propulsion Laboratory, California Institute of Technology

YUK L. YUNG

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125

*Submitted to the Astrophysical Journal*

## ABSTRACT

Spectra of exoplanet atmospheres provide us the opportunity to improve our understanding of these objects just as remote sensing in our own solar system has increased our understanding of the solar system bodies. The challenge is to quantitatively determine the range of temperatures and species abundances allowed by the data. This challenge is often difficult given the low information content of most exoplanet spectra which commonly leads to degeneracies in the interpretation. A variety of temperature and abundance retrieval approaches have been applied to exoplanet spectra, but no previous investigations have sought to compare these approaches. In this investigation we compare three different retrieval methods: Optimal Estimation, Differential Evolution Markov Chain Monte Carlo, and Bootstrap Monte Carlo. We call our suite of retrieval algorithms the CaltecH Inverse ModEling and Retrieval Algorithms (CHIMERA). We discuss what we can expect in terms of uncertainties in abundances and temperatures given current observations as well as potential future observations and what conclusions can be drawn given those uncertainties. In general we find that the three approaches agree for high quality spectra expected to come from potential future spaceborne missions, but disagree for low quality spectra representative of current observations. We also show that the Gaussian posterior probability distribution assumption made in the Optimal Estimation approach is valid for high quality spectral data. Furthermore we compare the results from a parameterized temperature profile versus a full classical Level-by-Level approach and discriminate in which situations each of these approaches is applicable. We also discuss the implications of our models for the inferred C to O ratios of exoplanetary atmospheres, which of course are important for understanding formation environments. More specifically we show that in the observational limit of a few photometric points, the retrieved C/O is biased towards values near solar and near one simply due to the assumption of uninformative priors.

*Subject headings:* planetary systems — planets and satellites: atmospheres — radiative transfer—

## 1. INTRODUCTION

Thermal emission spectra ( $\sim$  1-30 microns) of extrasolar planets can tell us about their atmospheric temperatures and compositions (see e.g., Tinetti et al. 2007; 2010; Grillmair et al. 2007; 2008; Swain et al. 2009a; 2009b; Madhusudhan & Seager 2009; Madhusudhan et al. 2011; Lee et al. 2012; Line et al. 2012). At the moment, these observations come in two types, broadband photometry mainly from the Spitzer Space Telescope (see e.g., Knutson et al. 2010) and ground-based instruments (Croll et al. 2010; Anderson et al. 2010; Gibson et al. 2010; Deming et al. 2012; Gillon et al. 2012), as well as higher resolution spectra such as Hubble Space Telescopes Wide Field Camera 3 (WFC3) (Berta et al. 2012; Swain et al. 2012; Deming et al. 2013) and Near Infrared Camera and Multi-Object Spectrometer (NICMOS) (Swain et al. 2009a; 2009b; Tinetti et al. 2010; Gibson et al. 2011; Crouzet et al. 2012). From these observations, signatures of a variety of molecules have been detected including  $H_2O$ ,  $CH_4$ , CO and  $CO_2$  (Tinetti et al. 2007; Swain et al. 2009a; 2009b; Tinetti et al. 2010), although the robustness of some of these detections have recently been called into question (Gibson et al. 2011). These same data have been used to infer the presence of atmospheric temperature inversions for a subset of hot Jupiters (e.g., Burrows et al. 2007; Knutson et al. 2008; 2010; Fortney et al. 2008; Madhusudhan & Seager 2009; 2010).

While the above studies have given us insight into the nature of these planetary atmospheres, very few have focused on the uncertainties in temperatures and compositions. Until relatively recently (Madhusudhan & Seager 2009, Madhusudhan et al. 2011, Lee et al. 2012, and Line et al. 2012), most compositions and temperatures and thus the subsequent conclusions, were determined through self-consistent forward modeling approaches that only explore a few potential solutions without a well-defined characterization of the uncertainty distributions of the physical parameters (e.g., Burrows et al. 2005; Fortney et al. 2005; Burrows et al. 2007). Furthermore, some self-consistent solutions make physical assumptions that may not necessarily be valid in exoplanetary atmospheres such as the assumption of thermochemical equilibrium gas concentrations or radiative-convective temperature structures (that is they may ignore other potentially important processes such as vertical mixing, photochemistry, zonal winds etc.). Additionally, this forward modeling approach is often unguided by the data and primarily driven by preconceived notions of how the atmosphere “should” look (as pointed out by Lee et al. 2012 and Benneke & Seager 2012) with the best solutions being the few that provide the lowest values of chi-squared.

In order to more rigorously characterize the ranges of allowable temperatures and compositions, Madhusudhan & Seager (2009) developed a multidimensional grid search approach which can fully characterize the uncertainty distributions for each parameter. Subsequent

studies (Madhusudhan et al. 2011; Benneke & Seager 2012) used the more sophisticated Markov Chain Monte Carlo approach (MCMC) to accomplish this goal. However, such approaches require the computation of many millions of models in order to fully characterize the parameter uncertainties which may not be feasible for more sophisticated forward models with many free parameters. In order to remedy this problem Lee et al. (2012) and Line et al. (2012) used the much faster Optimal Estimation (e.g, Rodgers 2000) approach to estimate the error distributions of each parameter. This approach is much faster due to the assumption that the parameter error distributions are Gaussian. However, this Gaussian assumption may result in an incorrect estimate of the error distributions (Benneke & Seager 2012).

The goals of this paper are to first understand the composition and temperature uncertainty distributions for different degrees of observational quality, and second to understand how those derived uncertainty distributions differ between the two fundamental parameter estimation approaches, Optimal Estimation and MCMC. This investigation represents the first direct comparison and synthesis of these retrieval approaches as applied to exoplanet atmospheres. A secondary goal is to understand how the derived composition uncertainties propagate into the C/O uncertainty. We accomplish these goals by comparing three different retrieval algorithms: Optimal Estimation (OE), a new MCMC algorithm known as Differential Evolution Markov Chain Monte Carlo (DEMC), and the model-dependent Bootstrap Monte-Carlo approach (BMC). This investigation is analogous to the investigation carried out by Ford (2005) on radial velocity data. First we will describe the three different retrieval techniques as well as our forward model in §2. We call our three-pronged retrieval approach CHIMERA-CaltechH Inverse ModEling and Retrieval Algorithms. Second, we compare the three spectral retrieval methods on different synthetic spectral data sets of varying observational quality in order to asses the robustness of the error estimations from each approach in §3. We will also compare the parameterized temperature profile approach (e.g., Madhusudhan & Seager 2009; Line et al. 2012) with the Level-by-Level profile approach (Lee et al. 2012). Finally, we will discuss the implications of these uncertainties for the estimated C to O ratios.

## 2. METHODS

In this section we describe the retrieval techniques, the forward model, and the parameterizations we use to retrieve the temperatures and compositions from thermal emission spectra.

### 2.1. The Retrieval Techniques

We use three different retrieval techniques to infer the compositions and temperatures from a spectrum. The techniques are inherently Bayesian as they attempt to solve the inverse problem by summarizing the full shape of the posterior in terms of the location in parameter space of the maximum likelihood and the uncertainties about that location. The first, and the fastest (least number of forward model calls) of these approaches, is

Optimal Estimation, the second is the model-dependent Bootstrap Monte Carlo, and the third is Differential Evolution Markov Chain Monte Carlo.

### 2.1.1. Optimal Estimation (OE)

The Optimal Estimation retrieval approach is well established in the fields of Earth atmosphere remote sensing (Rodgers 1976; Towney 1996; Rodgers 2000; Livesay et al. 2006; Kuai et al. 2013), solar system atmosphere remote sensing (Conrath et al. 1998; Irwin et al. 2008; Nixon et al. 2007; Fletcher et al. 2007; Greathouse et al. 2011), and recently exoplanet atmosphere remote sensing (Lee et al. 2012; Line et al. 2012). The basic approach is to minimize a cost function to obtain the *maximum a posteriori* (MAP) solution. Using Bayes theorem and the assumption that the data likelihood, the prior, and the posterior are Gaussian, one can derive the following cost function (or log likelihood):

$$\begin{aligned} J(\mathbf{x}) = & (\mathbf{y} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_e^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x})) \\ & + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) \end{aligned} \quad (1)$$

where  $\mathbf{y}$  is the set of  $n$  observations,  $\mathbf{x}$  is the set of  $m$  parameters which we wish to retrieve or the state vector,  $\mathbf{F}(\mathbf{x})$  is the forward model that maps the state vector onto the observations, and  $\mathbf{S}_e$  is the  $n \times n$  data error covariance matrix (typically off diagonal terms are zero and the diagonal elements are the square of the  $1\sigma$  errors of the observations).  $\mathbf{x}_a$  is the *a priori* state vector and  $\mathbf{S}_a$  is the  $m \times m$  *a priori* covariance matrix. The first term in equation 1 is simply “chi-squared” and the second term represents the prior knowledge of the parameter distribution before we make the observations. For high quality observations the second term is generally not important as most of the information in constraining the state vector comes from the observations. For low quality observations it is just the opposite. Following Irwin et al. 2008 we minimize equation 1 with Newton’s iteration method given by

$$\begin{aligned} \mathbf{x}_{i+1} = & \mathbf{x}_a + \mathbf{S}_a^{-1} \mathbf{K}_i^T (\mathbf{K}_i \mathbf{S}_a^{-1} \mathbf{K}_i^T + \mathbf{S}_e^{-1}) \\ & (\mathbf{F}(\mathbf{x}) - \mathbf{y} - \mathbf{K}_i(\mathbf{x}_a - \mathbf{x}_i)) \end{aligned} \quad (2)$$

where  $i$  is the iteration index and  $\mathbf{K}_i$  is the Jacobian matrix at  $i$  ( $K_{nm} = \frac{\partial F_n}{\partial x_m}$ ). Rather than taking the full Newton step we damp the solution with

$$\mathbf{x}'_{i+1} = \mathbf{x}_i + \frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{1 + \zeta} \quad (3)$$

where  $\zeta$  is the damping parameter. At each iteration we evaluate  $\mathbf{J}(\mathbf{x}_{i+1})$  and  $\mathbf{J}(\mathbf{x}'_{i+1})$ . If the latter is smaller, we set the state vector for the next iteration to  $\mathbf{x}'_{i+1}$  and decrease  $\zeta$  by 0.3. Otherwise we keep increasing  $\zeta$  by a factor of 10 and re-evaluate equation 3 and equation 1 until  $\mathbf{J}(\mathbf{x}'_{i+1})$  becomes less than  $\mathbf{J}(\mathbf{x}_{i+1})$ . Convergence is achieved when  $\mathbf{J}$  changes by less than  $1 \times 10^{-6}$  from the previous iteration, which typically occurs after  $\sim 10$ s of iterations. The resulting state vector is the MAP solution. The uncertainties on the state vector parameters are given by the posterior covariance matrix

$$\hat{\mathbf{S}} = (\mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \quad (4)$$

This matrix represents a multi-dimensional *normal* distribution (see Rodgers 2000 for the derivation). The diagonal elements are the square of the marginalized errors whereas the off diagonal terms describe the correlations/degenracies amongst the parameters. The first term  $, \mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K}$ , represents the uncertainties due to the measurement errors. This term uses the local gradient information to estimate the parameter uncertainties. The second term represents the prior uncertainties before making the measurements, which has less influence for higher quality data. Again, the major assumption in equations 1 and 4 is that the posterior for each parameter is Gaussian. This assumption is only valid when the region in phase space over which the forward model can be linearized is broader than the parameter uncertainties. However, it is this assumption that allows this approach to be extremely fast requiring only tens of forward model calls. As we shall see in §3, this assumption is valid for data that is of “good” quality, but breaks down for “poor” quality data.

### 2.1.2. Model Dependent Bootstrap Monte-Carlo (BMC)

A common way to more robustly characterize errors is through a Monte-Carlo resampling of the data (see e.g., Press et al. 2002 Chapter 15.6, Ford 2005 §4.2) in which many thousands of realizations of the original data (in our case, the spectra) are created using the uncertainties from the original dataset. These synthetic data are then refit using, say, OE, and the resulting best fit parameter distributions represent the uncertainties. There are multiple ways of generating the synthetic data realizations. The most common way is the residual resampling approach in which data realizations are created by adding the randomly shuffled residual between the best fit model and the data back to the original best fit model. This new realization is then fit and the process is repeated many times. The approach we take is similar, but rather than generate a new spectrum using the residual, we simply take the best fit, from OE, and then resample each point by drawing it from a normal distribution with a mean given by the best fit value and the width given by the data error bar for that point. We chose this approach over the residual resampling approach because sparse coverage spectra, like those from broadband observations, have virtually no residual as they can be fit perfectly due to the greater number of parameters than data points. We typically generate  $\sim 1000$  spectra realizations that are then refit by optimal estimation to obtain the state vector parameter distributions.

### 2.1.3. Differential Evolution Markov Chain Monte Carlo (DEMC)

The MCMC approach has revolutionized parameter estimation and error analysis in many fields. It is routinely used in radial velocity (Ford 2005 ) and transit light curve (e.g., Eastman, Gaudi, & Agol 2013) error analysis. Results from a well converged MCMC analysis can generally be considered as the best possible representation of the parameter uncertainties. Recently, this approach has been applied to the exoplanet atmosphere retrieval problem (Madhusudhan et al. 2011; Benneke & Seager 2012). Unlike optimal estimation, MCMC approaches make no assumptions about the shape of the posterior, but rather evaluate the posterior with millions of samples.

The basic approach of MCMC is to sample the posterior through a random walk process. The random walk is carried out by drawing states from some proposal distribution and evaluating whether or not the proposed state has an increased likelihood over the previous. Typically the proposal distribution is a normal distribution with a mean given by the current state in the chain ( $\mathbf{x}_i$ ) and a user defined width to achieve a particular acceptance rate (Gibbs sampling or Metropolis-Hastings). If the proposed state ( $\mathbf{x}_p$ ) has an improved likelihood over the current state, then that state is kept ( $\mathbf{x}_{i+1} = \mathbf{x}_p$ ) and a new proposal is made from that location. If the proposal state has not improved the likelihood, then that state is either rejected or accepted with some probability. This previous state dependent random walk constitutes a Markov Chain. Given enough samples this Makov Chain will converge to the target posterior (see Ford 2005 for a more detailed explanation).

Rather than standard MCMC approaches, we use an adaptive algorithm known as Differential Evolution Markov Chain Monte Carlo (DEMC) (ter Braak 2006; ter Braak & Vrugt 2008). The purpose of this approach is to obtain more appropriate proposal states by identifying the proper scale and orientation of the current estimate of the posterior. This scale and orientation information comes from the chain history. This approach gives a more efficient probing method for highly correlated parameter spaces and yields improved convergence rates. Our DEMC procedure is as follows:

1. Apply the OE technique to the measurements to obtain the best fit state vector and posterior covariance matrix,  $\hat{\mathbf{S}}$ . This step provides an initial estimate of the posterior.
2. Initialize  $N_{init}$  links ( $\mathbf{x}_{i=0-N_{init}}$ ) in each of  $N_{chains}$  (typically 3 chains, more chains will slow convergence) independent chains (arrays) by randomly drawing state vectors from the multivariate normal described by the posterior covariance matrix from step 1. Set the last link in one of the chains to the best fit state vector obtained in step 1. This step provides a good starting history from which our initial proposal states can be drawn. Combine each of the independent chains into one long chain that composes the history,  $\mathbf{X}_{history}$ .
3. Evaluate the cost function,  $\mathbf{J}$ , in equation 1 for the last link in each of the chains. If using a flat prior ignore the second term. This, again, is simply the equivalent of evaluating chi-squared.
4. Draw two random numbers,  $R_1$  and  $R_2$ , between zero and  $N_{chains} \times i$ , where  $i$  is the current state in the chain. Initially,  $i = N_{init}$ . Evaluate the proposed jump state given by

$$\mathbf{x}_p = \mathbf{x}_i + \gamma(\mathbf{x}_{R1} - \mathbf{x}_{R2}) + \mathbf{e} \quad (5)$$

where  $\mathbf{x}_{R1}$  and  $\mathbf{x}_{R2}$  are the states from different points in the the chain history,  $\mathbf{X}_{history}$ .  $\gamma$  is a scale factor typically set to  $2.38/\sqrt{2*m}$  (ter Braak 2006), where  $m$  is the number of parameters. This factor is meant to give acceptance rates of  $\sim 0.23$  for large  $m$ .  $\mathbf{e}$  is a vector drawn from a multivariate normal distribution

with a small variance relative to the chain variance in order to introduce a small amount of additional randomness. Repeat this process for the other  $N_{chains} - 1$  chains.

5. Evaluate the Metropolis (Metropolis et al. 1953) ratio,  $r = P(\mathbf{x}_p)/P(\mathbf{x}_i) = e^{-\frac{1}{2}(\mathbf{J}(\mathbf{x}_p)-\mathbf{J}(\mathbf{x}_i))}$ . If  $r$  is larger than 1, set  $\mathbf{x}_{i+1} = \mathbf{x}_p$  and if it is smaller only accept if it is larger than a random number between 1 and 0. Otherwise do not update the chain, set  $\mathbf{x}_{i+1} = \mathbf{x}_i$ . Repeat for the other  $N_{chains} - 1$  chains. Add the updated links in all  $N_{chains}$  to  $\mathbf{X}_{history}$ .

6. Repeat steps 4 and 5 until convergence is met. Convergence can be determined by looking at the trace plots of  $\mathbf{X}_{history}$  for each parameter or by using the Gelman-Rubin statistic on the set of  $N_{chains}$  chains. For this we use the algorithm from Eastman, Gaudi, & Agol 2013 which requires the Gelman-Rubin statistic to be less than 1.01 and the number of independent draws to be greater than 1000 for each parameter. Convergence typically occurs in less than  $10^5$  links in each of the  $N_{chains}$  for a total of  $N_{chains} \times 10^5$  links. This is about an order of magnitude less than parallel tempering or pure Metropolis-Hastings.

## 2.2. The Forward Model

The forward model,  $\mathbf{F}(\mathbf{x})$ , is the most important part of any retrieval algorithm. It is what maps the state vector of retrievable parameters onto the observations. In the case of atmospheric retrieval, the forward model takes temperatures and compositions and generates a model spectrum. Our particular forward model numerically solves the planet-parallel thermal infrared radiation problem for an absorbing, emitting atmosphere (we neglect scattering). We first divide the atmosphere into  $N_{lev}$  discretized atmospheric layers. The absorption optical depth for the  $k^{th}$  gas in the  $z^{th}$  layer at wavelength  $\lambda$  is

$$\Delta\tau_{k,z,\lambda} = f_{k,z}\sigma_{k,z,\lambda} \frac{\Delta P_z}{\mu_{atm}g} \quad (6)$$

where  $f_{k,z}$  is the volume mixing ratio of the  $k^{th}$  gas in the  $z^{th}$  layer,  $\sigma_{k,z,\lambda}$  is the absorption cross section per molecule of the  $k^{th}$  gas in  $z^{th}$  layer at wavelength  $\lambda$ ,  $\Delta P_z$  is pressure thickness of the  $z^{th}$  slab,  $\mu_{atm}$  is the mean molecular weight of the atmosphere and  $g$  is the gravity. The absorption cross sections are pre-computed on a  $1\text{ cm}^{-1}$  wavenumber grid at 20 evenly spaced temperature and log-pressure points from 500-3000K and  $50\text{-}10^{-6}$  bars respectively (similar to Sharp & Burrows 2007). The cross sections for each wavelength on the pre-computed grid are interpolated to the atmospheric temperatures and pressures in the  $z^{th}$  slab. To compute the total slab optical depth we sum the contribution from each gas to obtain

$$\Delta\tau_{z,\lambda} = \sum_{k=1}^{N_{gas}} \Delta\tau_{k,z,\lambda} \quad (7)$$

Upon computing the optical depths at each level, we can now solve for the upwelling irradiance with

$$I_\lambda = \sum_{z=0}^{N_{lev}} B_\lambda(T_z) e^{-\sum_{j=z}^{N_{lev}} \Delta\tau_{j,\lambda}} \Delta\tau_{z,\lambda} \quad (8)$$

where  $N_{lev}$  is the number of atmospheric levels and  $B_\lambda(T_z)$  is the Planck function at wavelength  $\lambda$  and temperature in the  $z^{th}$  slab. We use 90 atmospheric layers to compute the upwelling flux.

An important part of the forward model when using the Optimal Estimation approach is the computation of the Jacobian, or the sensitivity to the upwelling irradiance with respect to the desired retrievable parameters. When possible, it is preferable that the Jacobian be calculated analytically for both improvements in speed and in accuracy. We are interested in the retrieval of both abundances and temperatures so we must compute the Jacobian with respect to both the abundances and temperatures. We make the assumption of vertically uniform gas mixing ratios throughout the atmosphere and hence,  $f_{k,z}$  is independent of  $z$ . We now differentiate equation 8 with respect to the uniform gas mixing ratios for each gas  $f_k$  to obtain

$$\begin{aligned} \frac{\partial I_\lambda}{\partial f_k} = & \sum_{z=0}^{N_{lev}} B_\lambda(T_z) e^{-\sum_{j=z}^{N_{lev}} \Delta\tau_{j,\lambda}} \frac{\Delta\tau_{k,z,\lambda}}{f_k} \\ & - \sum_{z=0}^{N_{lev}} (B_\lambda(T_z) e^{-\sum_{j=z}^{N_{lev}} \Delta\tau_{j,\lambda}} \Delta\tau_{z,\lambda} \sum_{j=z}^{N_{lev}} \frac{\Delta\tau_{k,j,\lambda}}{f_k}) \quad (9) \end{aligned}$$

The first term is due to the changing emissivity of the emitting slab and the second term is how the change in transmittance affects the upwelling irradiance.

The sensitivity of the irradiance to a change in temperature in the  $z^{th}$  slab is given by

$$\frac{\partial I_\lambda}{\partial T_z} = (e^{-\sum_{j=z+1}^{N_{lev}} \Delta\tau_{j,\lambda}} - e^{-\sum_{j=z}^{N_{lev}} \Delta\tau_{j,\lambda}}) \frac{\partial B_\lambda(T_z)}{\partial T_z} \quad (10)$$

This equation is similar to equation 14 in Irwin et al. 2008 but we have neglected the first and last terms in their formula as they are small.

Since the observations are reported as the ratio of the planet flux to the stellar flux and not the irradiance, we perform a disk integration of equations 8-10 using four point Gaussian quadrature and then divide by an interpolated PHOENIX stellar flux grid model (Allard et al. 2000).

We include only CH<sub>4</sub>, CO<sub>2</sub>, CO, H<sub>2</sub>O, H<sub>2</sub>, and He in our model. H<sub>2</sub> and He are fixed in our models at thermochemically justifiable abundances. The exact abundances of these species is not critical as the sensitivity of the spectrum to H<sub>2</sub> and He is minimal. We retrieve only CH<sub>4</sub>, CO<sub>2</sub>, CO, and H<sub>2</sub>O. We choose these species because they are the most spectroscopically active and abundant species. Admittedly we could/should include every possible atmospheric constituent but this would be unwieldy and reliable high temperature absorption line lists only exist for a few. On that note, we use the HITEMP database (Rothman et al. 2010) to compute the tabulated cross-sections for CO<sub>2</sub>, CO, and H<sub>2</sub>O and the STDS database for CH<sub>4</sub> (Wenger & Champion 1998). Below 1.7  $\mu\text{m}$  for CH<sub>4</sub> we simply use the HITRAN (Rothman et al. 2009) database for lack of anything better (to the best of our knowledge). We use the Barysow et al. (2001;2002) and Jørgensen et al. (2000) databases for the computation of the H<sub>2</sub>-H<sub>2</sub>/He collision induced opacities. The Reference Forward Model (RFM-<http://www.atm.ox.ac.uk/RFM/>) was used to compute

the tabulated cross sections from the line strength databases. We have validated our forward model through a detailed comparison with the Oxford NEMESIS group (e.g., Lee et al. 2012) and our results agree to better than 5% (see Figure 1) (see online supplementary material).

An additional component of the forward model is the instrumental function used to convolve with the high-resolution model spectrum. For the broadband points we simply integrate the flux from the high-resolution model spectrum with the appropriate filter function for that point. When fitting higher resolution observations, the instrumental function is assumed to be a Gaussian (valid for grating spectrometers) in wavelength with a full width at half maximum (FWHM) determined by observations.

Now that we have a well defined forward model we can define our state vector. Again, we wish to retrieve the abundances of CH<sub>4</sub>, CO<sub>2</sub>, CO, and H<sub>2</sub>O and the temperature profile. More specifically, we choose to retrieve the log of the abundances as they can vary by orders of magnitude and to prevent negative mixing ratios. Our state vector is given by

$$\mathbf{x} = [\log(f_{H_2O}), \log(f_{CH_4}), \log(f_{CO}), \log(f_{CO_2}), T]^T. \quad (11)$$

where the  $f_k$ 's are all assumed constant with altitude. We feel this is appropriate for two reasons. First, vertical mixing will smooth out the mixing ratio profiles over the thermal infrared photosphere (Line et al. 2010, Moses et al. 2011, Line et al. 2011), and secondly current observations simply do not have the information content to warrant the retrieval of vertical mixing ratio information (see Lee et al. 2012). In the next section we describe how to go about retrieving the temperature profile.

### 2.2.1. Parameterized vs. Level-by-Level (Level-by-Level) Temperature Profile

We employ two approaches to retrieve the temperature profiles. The first, and the most commonly used in Earth and solar system atmosphere retrieval problems, is the Level-by-Level approach. This is the approach used in Lee et al. 2012. The second is a parameterized temperature profile approach similar to the approach used in Madhusudhan & Seager 2009 and Line et al. 2012. Each has its advantages and disadvantages described below.

The Level-by-Level temperature retrieval approach seeks an estimate of the temperature at each of the  $N_{lev}$  model layers. This approach is advantageous in that there are no pre-conceived assumptions made about how the atmospheric temperature should be parameterized. If the spectral signal-to-noise and resolution are high enough, there is generally enough sensitivity to obtain information about the temperature in individual atmospheric layers. However, there is a finite vertical resolution given the quality of the observations. Typically this resolution is set by the width of the thermal emission weighting functions and how much they overlap. Generally, when the spectra are noisy the Level-by-Level approach fits the noise which results in unphysical structure in the retrieved temperature profile. This is analogous to fitting a high-degree polynomial to only a few points. There are ways to smooth unphysical structure, one of them being to assume a correlation among the atmospheric layers (Rodgers 2000, Irwin 2008) implemented

through the prior covariance matrix,  $\mathbf{S}_a$ , with

$$S_{a,ij} = (S_{a,ii}S_{a,jj})^{1/2}e^{-\frac{|ln(P_i/P_j)|}{h}}. \quad (12)$$

Here  $P_i$  and  $P_j$  are the pressures at the  $i^{th}$  and  $j^{th}$  levels, respectively and  $h$  is the correlation length that controls the level of smoothing. The correlation length can be thought of as the number of scale heights over which the temperatures are correlated. For our simulations we choose  $h = 7$  as this provides a sufficient level of detail without producing unphysical oscillations. When using this approach our state vector is exactly as it is in equation 11 with  $T$  being an  $N_{lev}$  vector of temperatures at each level. The Level-by-Level approach is only appropriate when the information content of the spectra is sufficiently high such that the addition of the  $N_{lev}$  additional parameters is justified. For most current exoplanet spectra, this is an invalid approach.

The second temperature profile retrieval approach makes use of a parameterization. This approach is advantageous when the information content of a spectrum is low as the number of free variables is greatly reduced. However, the parameterization does force the retrieved atmospheric temperature structure to conform only to the profile shapes and physical approximations allowed by that parameterization. For our particular parameterization, we assume the atmosphere to be in radiative equilibrium based upon the analytic radiative equilibrium temperature profile of Guillot 2010 (and others such as Hansen 2008, Robinson & Catling 2012). This is the same parameterization used in Line et al. 2012. This profile assumes two independent downwelling visible channels of radiation and one upwelling stream of thermal emission. Briefly, the temperature as a function of the thermal optical depth,  $\tau$ , is given by

$$T^4(\tau) = \frac{3T_{int}^4}{4}\left(\frac{2}{3} + \tau\right) + \frac{3T_{irr}^4}{4}(1-\alpha)\xi_{\gamma_1}(\tau) + \frac{3T_{irr}^4}{4}\alpha\xi_{\gamma_2}(\tau) \quad (13)$$

with

$$\xi_{\gamma_i} = \frac{2}{3} + \frac{2}{3\gamma_i}[1 + (\frac{\gamma_i\tau}{2} - 1)e^{-\gamma_i\tau}] + \frac{2\gamma_i}{3}(1 - \frac{\tau^2}{2})E_2(\gamma_i\tau) \quad (14)$$

where  $\gamma_1$  and  $\gamma_2$  are the ratios of the Planck mean opacities in the visible streams to the thermal stream and the parameter  $\alpha$  (range 0 to 1) partitions the flux between the two visible streams.  $E_2(\gamma\tau)$  is the second order exponential integral function. The internal heat flux is parameterized by the temperature,  $T_{int}$ , which is fixed since it has little impact on the spectra. The stellar input at the top of the atmosphere is represented by  $T_{irr}$  given by

$$T_{irr} = \beta(\frac{R_*}{2a})^{1/2}T_* \quad (15)$$

where  $R_*$  and  $T_*$  are the stellar radius and temperature,  $a$  is the semi-major axis, and  $\beta$  is a catch all term on the order of unity for the albedo, emissivity, and day-night redistribution. The grey infrared optical depth can be mapped onto pressure coordinates using

$$\tau = \frac{\kappa_{IR}P}{g} \quad (16)$$

where  $P$  is the pressure,  $g$  the surface gravity (at 1 bar),

and  $\kappa_{IR}$  the Planck mean thermal infrared opacity. This  $\tau - P$  mapping assumes a linear relation between the optical depth and pressure, or a pressure independent infrared opacity. More complicated mappings that account for the pressure dependence of  $\kappa_{IR}$  can also be used (see e.g., Robinson & Catling 2012).

This temperature parameterization has 5 free parameters governing its structure:  $\kappa_{IR}$ ,  $\gamma_{v_1}$ ,  $\gamma_{v_2}$ ,  $\beta$ , and  $\alpha$ . Our parameterized state vector again, is given by equation 11 but with  $T$  replaced with  $[\kappa_{IR}, \gamma_{v_1}, \gamma_{v_2}, \beta, \alpha]$ . Combined with the gases this gives a total of 9 free parameters. The temperature *profiles* are then reconstructed from the probability distributions of those 5 parameters.

We should note that currently most exoplanet spectra often have fewer measurements than desired state variables. This means that each parameter cannot be uniquely determined. This is not a new problem (e.g., Madhusudhan & Seager 2009). This is why the prior is crucial. We can think of the prior as an “artificial” set of data on which the retrieval (all retrieval approaches) can rely when the measurements are insufficient to constrain a given parameter. Therefore the resulting constraints on a given parameter are a combination of the information obtained from the spectra and the prior knowledge. In the extreme case of no observational constraint, the posterior will simply be the prior. Hence, it is critical to choose an appropriate prior, especially for cases when there are more parameters than measurements.

With the optimal estimation formalism, we can assess the degree to which the constraint comes from the measured spectra versus the prior through what is called the averaging kernel. The averaging kernel is an  $m \times m$  matrix whose elements are given by

$$A_{ij} = \frac{\partial x_{i,retrieved}}{\partial x_{j,true}}. \quad (17)$$

where  $x_{i,retrieved}$  is the retrieved value of the  $i^{th}$  parameter and  $x_{j,true}$  is the true value of the  $j^{th}$  parameter. The diagonal elements tell us how much a retrieved parameter will respond to an actual change in that parameter in the atmosphere. For a given change in the true atmospheric state of some parameter  $i$ , if the measurements are perfect, we would expect to retrieve exactly that same change and hence, the value of  $A_{ii}$  would be one. If the measurements contribute nothing to our knowledge of parameter  $i$ , that is all of our knowledge of its value is from the prior, then  $A_{ii}$  will be zero. We can use this diagnostic to assess how heavily our error estimations are informed by the measurements. This is most important when using the Level-by-Level temperature profile retrieval. The sum of the diagonal elements of this matrix determines the total number of independent pieces of information that can be retrieved from the measurements. This can never be larger than the total number of individual data points.

### 3. TEST ON SYNTHETIC MEASUREMENTS

In this section, we apply the CHIMERA to a set of synthetic measurements in order to assess the robustness of each retrieval algorithm.

#### 3.1. Synthetic Observations

We create a generic hydrogen dominated hot jupiter planet and derive its emission spectrum in three different observing scenarios. Table 1 summarizes the basic planet parameters used to generate the model atmosphere and contrast spectrum. For simplicity we assume that the trace species have mixing ratios that are constant with altitude. Equations 13-16 are used to generate the atmospheric temperature profile of the planet from the values in Table 1. Figure 2 shows the model atmosphere and spectrum of the synthetic planet. The raw flux is divided by a PHOENIX stellar grid model that closely matches the chosen stellar properties. The thermal emission contribution functions (Figure 2, top-right) suggest that most of the emission originates between a few bars and a few mbars. Our synthetic data only provide believable estimates for the temperatures and abundances over this region of the atmosphere. The thermal contribution functions indicate that the emission from shorter wavelengths comes from deeper layers in the atmosphere and regions of high opacity tend to push the emission to higher altitudes. In this example, water is the dominant opacity source and acts almost like a continuum absorber across the spectrum (Figure 2, bottom-right). If we had no absorbing molecular species other than H<sub>2</sub>/He most emission would originate from the ~10 bar level.

We now create simulated observations for our synthetic planet under three different regimes. The first regime is a set of broadband observations through four of the Spitzer Infrared Array Camera (IRAC) channels at 3.6, 4.5, 5.7, and 8  $\mu\text{m}$  (top-left, Figure 3). This represents the spectral quality that is most commonly available for hot jupiters today. To create the synthetic observations, the spectrum in Figure 1 is first integrated over the IRAC filter functions at each channel and then random noise is added to each channel determined by the error bars size. The size of the error bars are representative of typical errors of IRAC observations (e.g., Machalek et al. 2009).

The second observational scenario is a multi-instrument case combining both Spitzer photometry, ground based photometry, and Hubble Wide Field Camera 3 spectra (WFC3) (top-right Figure 3). This combined set of observations from various instruments is more representative of the current level of observations that can be made today, and likely for the next half-decade, for many planets (e.g., WASP12b, WASP4b, HD209458b). Again, we use the same four Spitzer IRAC channels and error bars as before but also include ground based H and K<sub>s</sub> band photometry points. The error bars are taken from Crossfield et al. (2012). To create the synthetic WFC3 measurements (1.15-1.63  $\mu\text{m}$ ), the high resolution spectrum is convolved with a Gaussian instrumental profile with a FWHM of 0.0325  $\mu\text{m}$  with error bars taken from Swain et al. (2012). Random noise is again added to each point.

The third observational scenario illustrates the performance of a potential modest (by modest we mean reasonable cost) future spaceborne, FINESSE-like, telescope might produce (bottom, Figure 3). These simulated observations are created by convolving the high resolution spectrum with a moderate resolution Gaussian instrumental profile with a FWHM of 0.0075  $\mu\text{m}$  ( $R \sim 300$  at 2  $\mu\text{m}$ ). The measurement error bars, and hence random noise, are only suggestive and are based on a FINESSE-like noise model (Swain 2012). This spectral resolution is

comparable to that of Exoplanet Characterization Observatory (EchO) below 5  $\mu\text{m}$ , but less than the James Webb Space Telescope (JWST) Near Infrared Spectrometer.

Aside from the potential development of a ground-based near-IR spectroscopy program, most observations for the foreseeable future are likely to fall somewhere between the first and second cases. We are also being optimistic in our “worst” case observational scenario by including four broadband points instead of the now typical two from IRAC. In the latter case it is impossible to provide any unique constraints on the atmosphere without imposing many pre-conceived assumptions and priors.

### 3.2. The Prior

As mentioned in §2.1.1, the prior is important when the spectral information content is limited. We use a prior on both the gases and the temperature profile. For the purposes of this synthetic study, we assume Gaussian priors on the parameters that control the temperature profile and on the gas abundances. We could have chosen flat (un-informed) priors, however, the formalism of optimal estimation requires that the prior be Gaussian, and hence we maintain this prior for all of the retrieval approaches. We choose extremely broad Gaussian priors in order to mitigate the influence they have on the retrievals. For the temperature profile prior, we choose parameters that would reasonably match an atmosphere that is in radiative equilibrium over a wide range of conditions (e.g., variations in  $\kappa_{IR}$ ,  $\gamma_{v_1}$ ,  $\gamma_{v_2}$ ,  $\beta$ , and  $\alpha$ ). Table 2 lists the prior parameters we use in terms of the prior mean,  $\mathbf{x}_a$  and the prior covariance matrix,  $\mathbf{S}_a$ .

In addition to the Gaussian priors, we impose a lower limit on mixing ratios with a value  $1 \times 10^{-12}$  and an upper limit requiring the sum of the mixing ratios of the four retrieved gases to be less than 0.15. These limits attempt to bound what can be reasonably expected for the compositions of a hydrogen dominated atmosphere. Also, it would be impossible to detect a gas with an abundance less than 1 ppt in these simulations. We also impose a limit on the parameters that govern the temperature structure. We do not allow  $\kappa_{IR}$  to go above or below 10 and  $1 \times 10^{-4}$ , respectively. The lower limit is roughly the order of magnitude value of the Planck mean opacity expected for an all hydrogen atmosphere. The upper limit is a bit extreme but would be representative of an extremely opaque atmosphere. The upper and lower bounds on  $\gamma_1$  and  $\gamma_2$  are between 10 and  $1 \times 10^{-4}$  and are chosen to allow for a reasonable span of temperature profiles ranging from ones with inversions to ones nearly transparent to solar radiation.  $\alpha$  can only have physically meaningful values between 0 and 1.  $\beta$  cannot have values below 0, and we impose an artificial upper limit of 2. Generally these upper and lower limits rarely matter as most of the posteriors lie well within their ranges. Figure 4 shows the resulting temperature distribution and gas priors (inset). The prior temperature profile distributions are reconstructed by propagating the Gaussian prior probability distributions (including the above limits) of  $\kappa_{IR}$ ,  $\gamma_{v_1}$ ,  $\gamma_{v_2}$ ,  $\beta$ , and  $\alpha$  in Table 2 through equations 13-16. Upon reconstructing the temperature profiles there are thousands of temperatures for each pressure level. With these profiles a histogram of temperatures at each level can be constructed. Rather than show the “spaghetti diagram” with thousands of in-

dividual profiles, we show the  $1\sigma$  (68%) and  $2\sigma$  (95 %) confidence intervals at each pressure level. These confidence intervals are what is shown in Figure 4.

### 3.3. Results from the Parameterized Temperature Profile

We apply the three retrieval techniques to the three synthetic observations in Figure 3 under the radiative equilibrium temperature profile parameterization. This temperature profile has few parameters (5) because of the assumption of radiative equilibrium. The temperature prior in Figure 4 is used in all three techniques for all three retrieval cases. In each case, we retrieve the parameter distributions for the following state vector:

$$\mathbf{x} = [\log(f_{H_2O}), \log(f_{CH_4}), \log(f_{CO}), \log(f_{CO_2}), \log(\kappa_{IR}), \log(\gamma_{v_1}), \log(\gamma_{v_2}), \beta, \alpha]^T. \quad (18)$$

where again, the  $f_i$ s are the altitude independent volume-mixing ratios. We start by first applying the optimal estimation approach. In order to ensure that the retrieval does not get stuck in a local minimum, multiple starting guesses are used. These typically all converge to the same temperature and gas solution. As described in §2.1.3, the covariance matrix and the best fit from OE are then used to initialize the DEMC chains. Finally, the best fit from OE is used to initialize the synthetic measurement realizations used in the BMC. Figures 5-8 and Table 3 summarize the retrieval results and form the basis for the comparisons. The bounds quoted in Table 3 are for the 68% confidence intervals.

We must be careful in interpreting the confidence interval values when the posteriors extend to the imposed upper and lower limits, especially when those limits are somewhat arbitrary. Parameters with posteriors that approach the imposed lower limit will result in an overestimate of the lower bound on the confidence interval and an underestimate in the upper bound due to the imposed upper limit. In some cases if there were no Gaussian prior or no lower limit, the lower bound could extend to  $-\infty$ ! Of course we would interpret such a case as only having an upper bound.

Figure 5 shows the spectral fits. The first row shows the single best fit from optimal estimation. The second and third rows show the fits from the BMC and DEMC, respectively. Since the BMC and the DEMC provide many thousands of spectra, rather than plot each one, we summarize the fits by showing the median spectrum along with the  $1\sigma$ - and  $2\sigma$  spread at each wavelength. Illustrating the fits in this manner is more representative of the posterior than plotting spectra of different chi-squared levels. In other words, if a random set of parameters is drawn from the posterior, there would be a 95% chance that the flux at any one wavelength of the spectrum resulting from that parameter draw would fall within the  $2\sigma$  spread etc. There is little if any spread in the fits as the measurement quality improves. In the following subsections we summarize posteriors for the gas compositions, temperatures, and C to O ratios for each observational scenario.

#### 3.3.1. Gas Abundance Retrievals

Figures 6 and 7 summarize the gas mixing ratio retrieval results. Figure 6 shows histograms of the marginalized posterior for each of the four gases as a

result of each retrieval approach and observational scenario. We take the DEMC posterior (blue) to be representative of the true posterior. The optimal estimation posteriors (solid red curve) and the prior (dot-dashed red curve) are smooth because they are constructed analytically from the diagonal elements of  $\hat{\mathbf{S}}$ . Figure 7 is a “stair-step” plot that shows the correlations amongst the four gases comparing optimal estimation to the DEMC. For brevity, we do not show the Bootstrap Monte Carlo correlations. The solid blue filled regions are the  $1$  (dark) and  $2$  (light)  $\sigma$  confidence intervals derived from the DEMC, and the red curves are the  $1\sigma$  (inner) and  $2\sigma$  (outer) confidence intervals derived from OE.

The first column of Figure 6 shows the marginalized gas posteriors for the broadband observational scenario and the top set of panels of Figure 7 show the correlations amongst the gases. In this observational scenario, the three retrieval techniques produce quite different posteriors.  $H_2O$  has a fairly narrow posterior (relative to the prior) near the true state, suggesting that it is reasonably well constrained, at the  $1\sigma$  level, by even this low information content spectra. This is unsurprising as water is prevalent across all of the channels in this spectrum (see Jacobian in Figure 2). At the  $2\sigma$  level, however, the OE retrievals provide less of a constraint (Figure 7). The  $CH_4$  and  $CO_2$  posteriors abruptly fall off towards their upper end suggesting an upper bound constraint on these gases. The low end of their posteriors begin to track the prior down to the imposed lower limit indicating that from this observational scenario, there really is no observable lower limit to the abundances of these species.  $CH_4$  has a better defined upper edge than  $CO_2$  because both the  $3.6\ \mu m$  and  $8\ \mu m$  channels overlap with the strongest methane absorption bands.  $CO$  is virtually unconstrained by the synthetic broadband measurements as it closely matches the prior across the full range of values. The difficulty in constraining  $CO$  and  $CO_2$  is due to the inability of the  $4.5\ \mu m$  broadband photometric measurement to decouple the  $CO$  and  $CO_2$  strong bands. There is a slight hint of an inverse correlation, as expected, from the optimal estimation results in Figure 7, however this correlation is not apparent from the DEMC results.

Our implementation of optimal estimation struggles to appropriately capture the errors in this observational scenario. This is because it approximates the posteriors with broad Gaussians which simply do not capture the appropriate structure. It does, however, do a fairly good job of determining the true state. The Gaussian approximation cannot appropriately handle upper bounds on  $CH_4$  and  $CO_2$ , causing an overestimate of the  $1\sigma$  upper bound. We note, that at least it is overestimating the errors rather than underestimating them. OE does a fine job at approximating the posterior for  $CO$ , which happens to be similar to the prior, again suggesting no constraint. This is reaffirmed by looking at the averaging kernel element,  $A$ , in Table 3 which shows that most (70%) of the information in constraining  $CO$  comes from the prior (see §2.2.1). The Bootstrap Monte Carlo (green curve) completely fails to appropriately capture the posterior in this particular observational scenario. This is because the different realizations from the BMC approach produce parameter distributions that are con-

fined to a small area of phase space localized near the best fit solution from optimal estimation and cannot therefore sample the entirety of the posterior.

The second column of Figure 6 shows the gas posteriors for the multi-instrument observational scenario. The information gain from this synthetic observational scenario is only marginal relative to the broadband case. Water has the largest improvement in uncertainty due to the leverage provided by the WFC3 spectra which covers the  $1.15\ \mu\text{m}$  and  $1.4\ \mu\text{m}$  water bands. Upon inspecting Figure 7 we find that the WFC3 data combined with the ground based photometric points trims the  $2\sigma$  tail but does little to improve the  $1\sigma$  uncertainties. Sadly, there is virtually no reduction in the uncertainties on  $\text{CH}_4$ ,  $\text{CO}$ , and  $\text{CO}_2$ . In fact the marginalized posteriors (Figure 6) produced by DEMC in this observational scenario look nearly identical to the previous case. These results are not surprising given that these species do not have strong features at the near-IR wavelengths of the WFC3 spectroscopy and the H and  $K_s$  photometry. Optimal estimation provides an accurate error estimation for water but appears to provide an overly optimistic estimation of the uncertainties on  $\text{CH}_4$ . OE, as in the previous scenario, captures the essence of large uncertainties on  $\text{CO}$  and  $\text{CO}_2$  with broad Gaussians. Bootstrap Monte Carlo underestimates the uncertainties in all species with the exception of water.

Finally, results for a hypothetical future spaceborne telescope are shown in the last column of Figure 6 and bottom set of panels of Figure 7. The reduction in uncertainties are staggering when compared to the previous observational scenarios. All of the gases are constrained to within better than an order of magnitude. This is an orders-of-magnitude improvement over the previous cases. The high signal-to-noise and high spectral resolution combine to provide excellent coverage of each of the four gases. With this high quality spectrum, all three retrieval approaches give the same results. The Gaussian assumption used in optimal estimation is perfectly appropriate in this case. The differences in the  $1\sigma$  uncertainties derived from OE are less than  $\sim 10\%$  than the uncertainties derived from DEMC. The two-dimensional confidence intervals in Figure 7 also agree quite well. The prior also plays very little role in the retrieval as shown with the near unit averaging kernel elements in Table 3.

### 3.3.2. Temperature Profile Retrievals

Figure 8 shows the temperature profile retrieval results ( $1\sigma$  (dark red) and  $2\sigma$  (light red)) reconstructed from the posteriors of  $\log(\kappa_{IR})$ ,  $\log(\gamma_{v_1})$ ,  $\log(\gamma_{v_2})$ ,  $f$ , and  $\alpha$ , as in Figure 4. The blue curve is the statistical median of these profiles. This median profile not representative of any one given temperature profile, and in fact this median profile may not even provide a good fit to the observations or adhere to the parameterization, but is shown simply as a statistical summary of all possible temperature profiles. The black curve is the true temperature profile.

The temperature profile posteriors for the broadband scenario (first column Figure 8) from OE and DEMC have similar widths and both capture the entire true temperature profile within the  $2\sigma$  interval. There is also a non-negligible ( $\sim 30\text{-}50\%$ ) reduction in the temperature precision compared with the prior, over the atmo-

spheric region probed by these observations. Outside of the range spanned by the thermal emission contribution functions (Figure 2), the temperature uncertainty grows and begins to relax back towards the prior where there is no observational constraint. Again, the BMC approach completely underestimates the error when compared with the other two approaches because of its inability to fully characterize the posterior outside of a small region of phase space localized around the OE original best fit.

Moving onto the multi-instrument observations (middle column Figure 8) we find a 22% reduction in the temperature uncertainty between 1 and 0.01 bars. The large number of spectral channels from the WFC3 data that have weighting functions over this region are the primary contributors to this increased precision. OE and BMC underestimate the temperature uncertainties relative to DEMC at 100 mbars, but the OE and DEMC have reasonable agreement over the entire profile.

The future spaceborne telescope observations improve the temperature uncertainties by a remarkable factor of  $\sim 4.5$  over the previous case. Optimal estimation slightly overestimates the temperatures outside the atmospheric levels probed by the observations. This overestimate is due to the overestimation of the  $\gamma_{v_1}$  and  $\gamma_{v_2}$  posterior widths. These two parameters control the relative difference between the upper atmosphere and lower atmosphere temperatures. Hence, extreme values of  $\gamma_{v_1}$  and  $\gamma_{v_2}$  will effect these regions more than the middle atmosphere. This why the OE and the DEMC temperatures agree in the middle atmosphere but not outside of it.

### 3.3.3. C to O Ratios

Determination of the C to O ratios of exoplanetary atmospheres is critical to the understanding of their atmospheric chemistry (Lodders & Fegley 2002; Moses et al. 2012) and formation environments (Öberg, Murray-Clay, & Bergin 2011). Given the abundance posteriors derived with CHIMERA, we can compute C/O posterior distributions (Figure 9). The C/O is calculated with the following formula,

$$\text{C}/\text{O} = \frac{\Sigma C}{\Sigma O} \approx \frac{\text{CH}_4 + \text{CO} + \text{CO}_2}{\text{H}_2\text{O} + \text{CO} + 2\text{CO}_2}. \quad (19)$$

There are a few simple things to note about this equation. When CO is the dominant species, C/O is 1. If  $\text{CO}_2$  is the dominant species, C/O will be  $1/2$ . When methane dominates the C/O will be large and when  $\text{H}_2\text{O}$  dominates C/O will be small. The solar C/O is 0.55. A number of exoplanets have reported C/O's near 1. We can construct the C/O probability distributions by propagating the probability distributions of each gas through equation 19, similar to the method used to construct the temperature profile posteriors. Before inspecting the posteriors derived from CHIMERA, we find it illustrative to investigate the prior. Upon propagating the Gaussian priors (with the limits) of the gases through equation 19 we obtain the C/O prior in Figure 9. We find that this prior has two peaks, one is at a C/O of 1 and the other is at a C/O of 0.5. The locations of these peaks are insensitive to whether or not the gas abundance priors are uniform or broad Gaussians. These peaks are also insensitive to the lower and upper bounds placed on either a

uniform or Gaussian gas prior. These double peaks are due to an elegant mathematical misfortune. For illustrative purposes, let us assume we draw the set of four gases from a uniform log prior. We would expect then, that one gas will have a larger abundance than the other three 25% of the time. This means that in roughly 25% of the draws CO will dominate, which would cause the ratio in equation 19 to be one, 25% of the time. A similar argument goes for CO<sub>2</sub> resulting in a C/O of 0.5 roughly 25% of the time. So we see, that if we did not observe a particular planet, and assumed uninformed priors on each of the gases, we would naturally conclude that the planet has equal chances of having a C/O of one or one-half. However, our priors are not uniform, but rather broad Gaussians, but they are broad enough that this behavior still occurs, with a slight preference for a C/O of one.

If we observe this planet with the four broadband points we obtain the posteriors in the upper-left panel of Figure 9. OE and DEMC produce similar C/O posteriors both of which maintain the two peaked features at 1 and 0.5 but with overall less power in the peaks. These features persist simply because the gas posteriors from DEMC and OE do not deviate too strongly from the prior. There is more probability in the lower C/O tail than in the prior because of the higher values of H<sub>2</sub>O preferred by the measurements over the prior. This is good, since the C/O for our fictitious planet is much less than one or one-half. BMC anomalously captures the true C/O at the peak of its posterior. Again, BMC greatly underestimates the posterior widths because it only searches a localized region about the OE best fit. Since the OE best fit gas abundances are very near truth (Figure 6), the BMC posteriors, which are highly localized about the OE best fit parameters, will overemphasize the C/O derived by that best fit.

The story is the same for the multi-instrument observational scenario. Unfortunately, at least in this example, it appears that the WFC3 and ground based data provide very little additional constraints in reducing the C/O uncertainty, with the double peaked feature from the prior persisting in the DEMC results.

Improving the observational quality further with a future spaceborne telescope essentially obtains the correct value to high precision. The peak of the posterior is far enough away from the double peaked prior that the results appear to be less influenced by the prior than the previous cases. All three retrieval approaches give a nearly identical posterior.

From this exercise we have learned that it is difficult to constrain the C to O ratio of an exoplanet atmosphere. Simple, uninformed, or nearly uninformed priors on the gas abundances produce a double peaked C/O prior at near solar value and one. Even in the best cases, current observations are likely to provide only an upper limit on this quantity. This result suggests that previously published claims to detect enhanced C to O ratios with photometry alone may be influenced by these subtle biases, and should be viewed with strong skepticism. We will discuss this issue in the context of specific exoplanets in a companion paper.

### 3.4. Results from the Level-by-Level Temperature Profile

The Level-by-Level temperature profile approach attempts to determine the temperature for each model

layer. In contrast to the parameterized temperature profile which had only 5 parameters, the Level-by-Level approach requires as many parameters as model layers, for a total of 90 parameters. This larger number of parameters is far greater than the number of meaningful constraints provided by most observations. However, this approach makes no assumptions about the physical structure of the temperature profile (e.g., radiative, radiative-convective, advection etc.). While there is no potentially biasing parameterization, the retrievals can result in unphysical temperature profiles. Obviously, the temperature at each of the 90 levels cannot be perfectly retrieved, but rather the retrievals have to depend on the prior when spectral information on the temperature is sparse.

For the Level-by-Level prior we assume an a prior variance of 400 K and covariance amongst each level with all other levels given by equation 12. The 400 K variance is used to produce a similar temperature profile prior as in Figure 4. This correlation helps reduce the effective number of levels that have to be independently retrieved. Admittedly, the degree of correlation is somewhat of an external arbitrary parameter, but it is chosen to avoid over-fitting (i.e., fitting to the noise) without hindering the Level-by-Level flexibility. It can be thought of as a smoothing, or more specifically, a regularization. We can also use the averaging kernel profile to assess where the temperature is constrained by the measurements versus the prior. The gas priors are the same as before. We choose only to compare the results from optimal estimation and the Bootstrap Monte Carlo. We do not attempt the DEMC approach on such a large ( $\sim 100$ ) number of parameters, as MCMC algorithms (to the best of our knowledge) are not well suited for large numbers of parameters because of the large number of steps required to fully map the  $n$ -dimensional probability distribution when  $n$  is large.

Figure 10 shows the spectral fits as a result of OE and BMC using the Level-by-Level temperature approach, similar to Figure 5. Figure 11 shows the marginalized gas posteriors. We find the gas posteriors and the agreement amongst the retrieval techniques are very similar to those derived in Figure 6 using the parameterized temperature profile. This is somewhat surprising given the extremely different temperature profile retrieval approaches. This suggests that the gas abundances can be properly and consistently retrieved regardless of the temperature profile assumptions. We could, however, imagine a contrived example in which the true temperature profile is so wildly different from what can be reasonably approximated with the parameterization, that the two approaches would yield differing gas posteriors.

Figure 12 shows temperature profile posteriors under the Level-by-Level temperature profile assumption. For the broadband scenario, the uncertainties more or less do not improve much beyond the prior. The greatest gain in improvement is over the region spanned by the averaging kernel (green curve). The uncertainty reduces from the prior uncertainty of  $\pm 400$  K to  $\pm 265$  K at 100 mbars. The BMC approach using the Level-by-Level temperature profile produces a much smaller error than the OE approach, and at some levels the  $2\sigma$  uncertainties do not even capture the true state. Again, this because the BMC is only able to characterize a highly localized region around the OE best fit.

The reduction in temperature uncertainty due to the addition of the WFC3 and ground-based photometry data is more apparent with the Level-by-Level approach than with the parameterized approach. The uncertainties in temperature at 100 mbars are reduced to  $\pm 177$  K, though smaller uncertainties are achieved at deeper levels due to the addition of the WFC3 data which probe deeper atmospheric levels. This is why the averaging kernel profile peaks at a deeper level. The BMC results show a larger uncertainty than they do in the broadband observational scenario but still greatly underestimate the profile spread relative to optimal estimation.

The future spaceborne telescope observations reduce the temperature uncertainties to  $\pm 70$  K, a factor of nearly four better than what can be done with the broadband observations. Outside of the region spanned by the averaging kernel uncertainties relax back to the prior widths. As before, in both cases the BMC approach underestimates the temperature uncertainties relative to the OE derived uncertainties. The uncertainties in temperature derived using the Level-by-Level temperatures are a factor of two larger than those derived with the parameterized temperature profile. This is because the retrievals with the parameterization only allow temperature profiles that conform to radiative equilibrium whereas the Level-by-Level retrievals can allow for a wider range of possibilities that do not necessarily have to conform to this constraint.

Another way to determine the robustness of the Level-by-Level retrieval is to explore the role of the prior temperature profile (see e.g., Lee et al. 2012). For this, we investigate the effect of three different temperature priors (different prior profiles,  $\mathbf{x}_a$  but same widths,  $\mathbf{S}_a$ ) on the retrieved profiles and check to see if they are consistent with the estimated errors (Figure 13). The shaded grey region in Figure 13 is the  $1\sigma$  retrieval uncertainty from Figure 12 using the nominal prior. Two of the other priors are the nominal profile with a  $\pm 500$  K offset, and the third is an isothermal profile set to the equilibrium temperature of the planet. In all three cases we find that the retrieved profiles fall within the  $1\sigma$  bounds of the nominal retrieved profile. This suggests that although different temperature profile priors are used, they generally produce retrieved profiles that are statistically consistent with each other. As the spectral quality improves, the different priors produce identically the same retrieved profiles over the atmospheric regions spanned by the thermal emission weighting functions. Outside of this region, the profiles diverge and relax towards their respective priors with no consequence on the spectra. This is yet another demonstration that the spectra are only sensitive to a small region of the atmosphere between a few bars up to a few mbars. While some of these Level-by-Level profiles may not be physical, especially in the broadband observational scenario, they are a more direct reflection of the information provided by the measurements in the absence of a parameterized model.

#### 4. DISCUSSION & CONCLUSIONS

We have developed a new statistically robust suite of exoplanet atmospheric retrieval algorithms known as CHIMERA. This suite consists of the Optimal Estimation (OE, §2.1.1), Bootstrap Monte Carlo (BMC, §2.1.2), and Differential Evolution Markov Chain Monte

Carlo (DEMC, §2.1.3) approaches and a validated forward model (§2.2). We have tested each of these approaches on the dayside thermal emission spectra for a synthetic planet under a variety of observational scenarios ranging from current observations to potential future observations (§3). In general, we find that the three retrieval approaches produce similar posteriors when the measurement quality is good, typically when there are more observed spectral channels than retrievable parameters (Figures 6 and 7). The Gaussian approximation made by Optimal Estimation breaks down for low quality measurements, but is perfectly valid for high quality measurements likely to come from future spaceborne observations. It also appropriately captures the correlations amongst the various parameters. This approach is much less of a computational burden than Monte Carlo approaches and will prove useful for quick reductions of large, high-quality data sets. The Optimal Estimation formalism also allows for the calculation of the averaging kernel (§2.2.1), which is a useful diagnostic to determine how much of the posterior is influenced by the prior versus the measurements. The Bootstrap Monte Carlo approach generally fails to capture the essence of the posteriors. This is because the regeneration of synthetic measurement realizations based on the Optimal Estimation best fit only sample a localized region of phase space near the best fit solution. This is especially problematic in the cases where there are fewer spectral data points than parameters. In this scenario, even with small measurement error, there will still be many possible best-fit solutions, thus creating enormous degeneracies among the parameters. Since the BMC is initialized with only one possible best-fit set of parameters out of many, the derived parameter uncertainties will only be representative of the uncertainties about that localized best fit. We therefore strongly advise against the Bootstrap Monte Carlo approach when the number of parameters is larger than the number of spectral data points. In the high signal-to-noise and high spectral resolution regime, both the BMC and OE methods provide reasonable parameter uncertainties. We have also introduced the application of Differential Evolution Markov Chain Monte Carlo to the spectral retrieval problem and found that convergence can be obtained efficiently by using an appropriate proposal distribution based on the chain history.

We find that for broadband observations typical of most hot Jupiters, very limited constraints on the gas abundances can be obtained. The Spitzer photometry does a particularly poor job constraining the relative abundance of CO, with most posteriors simply reflecting their priors (§3.3.1). The addition of WFC3 observations provide little additional constraint on the gas abundances derived from dayside thermal emission spectra, with the exception of a slight improvement on the water abundances. This is primarily due to the limited spectral coverage provided by the red grism on WFC3, which spans the wavelengths from 1.2-1.6  $\mu$ m. As the measurement quality improves, the parameter uncertainties decrease and become more Gaussian. Moderate cost future spaceborne instruments have the capability of obtaining better than order-of-magnitude constraints on gas compositions with their posteriors generally being independent of the prior. This is typically many orders of magnitude better than current observational capabilities (Table 3). The

derived gas posteriors are also independent of whether or not a parameterized or Level-by-Level temperature profile is used (Figure 6 vs. Figure 11).

Constraining the C to O ratio of exoplanet atmospheres is very difficult due to the broad nature of some of the gas posteriors, especially CO. In the absence of valid observational constraints the posteriors for these molecules simply reflect the priors, which produce a double-peaked distribution with maxima at C to O ratios of 0.5 and 1 (§3.3.3, Figure 9). Only high quality observations from the future spaceborne telescope scenario are independent of the double-peaked prior. As a result, caution must be taken when interpreting C to O ratios from broad gas posteriors.

Reasonable temperature constraints could be obtained in all observational scenarios and temperature retrieval approaches, though the Bootstrap Monte Carlo approach again fails to fully capture the posterior (§3.3.2, Figure 8). The temperature profiles and corresponding uncertainties can only be trusted for the region over which the thermal emission contribution functions peak, typically between a few bars and a few mbars (Figure 2). Outside of this window, the temperature profiles are strongly affected by their priors. The Level-by-Level temperature profile approach overestimates the temperature uncertainties compared with the parameterization due to the allowance of more profiles (§3.4). These Level-by-Level profiles can be unphysical but are more reflective of the measurements without imposing preconceived notions of how the physical structure of the atmosphere should behave. While this approach produces statisti-

cally consistent profiles in low quality observational scenarios, we would still recommend using a parameterization for said cases. However, for high quality spectra the Level-by-Level approach is recommended given its slightly more pessimistic temperature uncertainties and its non-dependence on a particular parameterization.

In a follow up investigation we will use CHIMERA to perform a uniform analysis of an ensemble of secondary eclipse spectra. Such a study will allow us to determine the biases introduced by the choice of fitting method for individual planets and to derive a uniform set of relative abundances and temperatures for these planets that can be reliably inter-compared and trends identified. This kind of uniform analysis has the potential to provide invaluable insights into exoplanetary atmospheric processes and formation environments.

## 5. ACKNOWLEDGEMENTS

We thank Jaimin Lee and Leigh Fletcher for their willingness to compare radiative transfer codes. We also thank John Johnson and Jonathan Fortney for useful conversations. We thank members of Yuk Yungs group for useful comments. This research was supported in part by an NAI Virtual Planetary Laboratory grant from the University of Washington to the Jet Propulsion Laboratory and California Institute of Technology. Part of the research described here was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. YLY was supported in part by NASA NNX09AB72G grant to the California Institute of Technology.

## REFERENCES

Allard, F., Hauschildt, P. H., & Schweitzer, A. 2000, ApJ, 539, 366  
 Anderson, D. R., Gillon, M., Maxted, P. F. L., et al. 2010, VizieR Online Data Catalog, 351, 39003  
 Benneke, B., & Seager, S. 2012, ApJ, 753, 100  
 Berta, Z. K., Charbonneau, D., Désert, J.-M., et al. 2012, ApJ, 747, 35  
 Borysow, A. 2002, A&A, 390, 779  
 Borysow, A., Jørgensen, U.G., Fu, Y. 2001, J. Quant. Spec. Radiat. Transf., 68, 235  
 Burrows, A., Hubeny, I., Budaj, J., Knutson, H. A., & Charbonneau, D. 2007, ApJ, 668, L171  
 Burrows, A., Hubeny, I., & Sudarsky, D. 2005, ApJ, 625, L135  
 Charbonneau, D., Allen, L. E., Megeath, S. T., et al. 2005, ApJ, 626, 523  
 Charbonneau, D., Knutson, H. A., Barman, T., et al. 2008, ApJ, 686, 1341  
 Conrath, B. J., Gierasch, P. J., & Ustinov, E. A. 1998, Icarus, 135, 501  
 Crisp, D., Atlas, R. M., Breon, F. M., Brown, L. R., Burrows, J. P., Ciais, P., Schroll, S. 2004, Adv. in Space Res., 34, 700-709.  
 Croll, B., Jayawardhana, R., Fortney, J. J., Lafrenière, D., & Albert, L. 2010, ApJ, 718, 920  
 Crossfield, I. J. M., Knutson, H., Fortney, J., et al. 2012, ApJ, 752, 81  
 Crouzet, N., McCullough, P. R., Burke, C., & Long, D. 2012, ApJ, 761, 7  
 Deming, D., Fraine, J. D., Sada, P. V., et al. 2012, ApJ, 754, 106  
 Deming, D., Wilkins, A., McCullough, P., et al. 2013, arXiv:1302.1141  
 Désert, J.-M., Sing, D., Vidal-Madjar, A., et al. 2011, A&A, 526, A12  
 Désert, J.-M., Bean, J., Miller-Ricci Kempton, E., et al. 2011, ApJ, 731, L40  
 Eastman, J., Gaudi, B. S., & Agol, E. 2013, PASP, 125, 83  
 Ford, E. B. 2005, AJ, 129, 1706  
 Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D., & Freedman, R. 2005, ApJ, 627, L69  
 Fletcher, L. N., Irwin, P. G. J., Teanby, N. A., et al. 2007, Icarus, 189, 457  
 Gibson, N. P., Aigrain, S., Pollacco, D. L., et al. 2010, MNRAS, 404, L114  
 Gibson, N. P., Pont, F., & Aigrain, S. 2011, MNRAS, 411, 2199  
 Gillon, M., Triaud, A. H. M. J., Fortney, J. J., et al. 2012, A&A, 542, A4  
 Greathouse, T. K., Richter, M., Lacy, J., et al. 2011, Icarus, 214, 606  
 Grillmair, C. J., Charbonneau, D., Burrows, A., et al. 2007, ApJ, 658, L115  
 Grillmair, C. J., Burrows, A., Charbonneau, D., et al. 2008, Nature, 456, 767  
 Guillot, T. 2010, A&A, 520, A27  
 Harrington, J., Hansen, B. M., Luszcz, S. H., et al. 2006, Science, 314, 623  
 Harrington, J., Luszcz, S., Seager, S., Deming, D., & Richardson, L. J. 2007, Nature, 447, 691  
 Hansen, B. M. S. 2008, ApJS, 179, 484  
 Irwin, P. G. J., Teanby, N. A., de Kok, R., et al. 2008, J. Quant. Spec. Radiat. Transf., 109, 1136  
 Jørgensen, U. G., Hammer, D., Borysow, A., & Falkesgaard, J. 2000, A&A, 361, 283  
 Knutson, H. A., Charbonneau, D., Allen, L. E., et al. 2007, Nature, 447, 183  
 Knutson, H. A., Charbonneau, D., Allen, L. E., Burrows, A., & Megeath, S. T. 2008, ApJ, 673, 526  
 Knutson, H. A., Lewis, N., Fortney, J. J., et al. 2012, ApJ, 754, 22  
 Lee, J.-M., Fletcher, L. N., & Irwin, P. G. J. 2012, MNRAS, 420, 170  
 Line, M. R., Liang, M. C., & Yung, Y. L. 2010, ApJ, 717, 496

Line, M. R., Vasisht, G., Chen, P., Angerhausen, D., & Yung, Y. L. 2011a, ApJ, 738, 32

Line, M. R., Zhang, X., Vasisht, G., et al. 2012, ApJ, 749, 93

Lodders, K., & Fegley, B. 2002, Icarus, 155, 393

Madhusudhan, N., & Seager, S. 2009, ApJ, 707, 24

Madhusudhan, N., Harrington, J., Stevenson, K. B., et al. 2011, Nature, 469, 64

Machalek, P., McCullough, P. R., Burrows, A., et al. 2009, ApJ, 701, 514

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953, The journal of chemical physics, 21, 1087

Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, ApJ, 737, 15

Nixon, C. A., Achterberg, R. K., Conrath, B. J., et al. 2007, Icarus, 188, 47

Öberg, K. I., Murray-Clay, R., & Bergin, E. A. 2011, ApJ, 743, L16

Press, W.H., Teukolsky, S., Vetterling, W.T., Fannery, B. 1995, Numerical Recipies: The art of scientific computing, sec. ed., Camb. Univ. Press

Robinson, T. D., & Catling, D. C. 2012, ApJ, 757, 104

Rodgers, C. D. 1976, Reviews of Geophysics and Space Physics, 14, 609

Rodgers, C.D., Inverse methods for atmospheric sounding, Theory and Practice, 2000

Rothman, L. S., Gordon, I. E., Barbe, A., et al. 2009, J. Quant. Spec. Radiat. Transf., 110, 533

Rothman, L. S., Gordon, I. E., Barber, R. J., et al. 2010, J. Quant. Spec. Radiat. Transf., 111, 2139

Stevenson, K. B., et al. 2010, Nature, 464, 1161

Swain, M. R., Vasisht, G., Tinetti, G., et al. 2009a, ApJ, 690, L114

Swain, M. R., Tinetti, G., Vasisht, G., et al. 2009b, ApJ, 704, 1616

Swain, M., Deroo, P., Tinetti, G., et al. 2012, arXiv:1205.4736

Swain, M. R. 2012, American Astronomical Society Meeting Abstracts #220, 220, #505.05

ter Braak, C. J. 2006, Statistics and Computing, 16, 239-249

ter Braak, C. J., & Vrugt, J. A. 2008, Statistics and Computing, 18, 435-446

Tinetti, G., Vidal-Madjar, A., Liang, M.-C., et al. 2007, Nature, 448, 169

Tinetti, G., Griffith, C. A., Swain, M. R., et al. 2010, Faraday Discussions, 147, 369

Tinetti, G., Deroo, P., Swain, M. R., et al. 2010, ApJ, 712, L139

Twomey, S., Herman, B., & Rabinoff, R. 1977, JAS, 34, 1085

Wenger, C., & Champion, J. P. 1998, J. Quant. Spec. Radiat. Transf., 59, 471

TABLE 1

PARAMETERS USED TO GENERATE THE FICTITIOUS MODEL ATMOSPHERE AND SPECTRUM.  $R_p$  IS THE PLANET RADIUS IN UNITS OF JUPITER RADII,  $R_{star}$  IS THE STELLAR RADIUS IN UNITS OF SOLAR RADII,  $T_{star}$  IS THE STELLAR EFFECTIVE TEMPERATURE,  $a$  IS THE SEMIMAJOR AXIS,  $T_{int}$  IS THE INTERNAL HEAT FLUX TEMPERATURE OF THE PLANET,  $g$  IS THE PLANETARY SURFACE GRAVITY.  $\gamma_{v1}$ ,  $\gamma_{v2}$ ,  $\kappa_{IR}$ ,  $\alpha$ , AND  $\beta$  ARE THE PARAMETERS THAT CONTROL THE SHAPE OF THE RADIATIVE EQUILIBRIUM TEMPERATURE PROFILE. THE  $f_i$ 'S ARE THE CONSTANT-WITH-ALTITUDE VOLUME MIXING RATIOS FOR EACH SPECIES IN PARTS-PER-MILLION (PPM).

Parameter	Value
$R_p(R_J)$	1.138
$R_{star}(R_{sun})$	0.756
$T_{star}(K)$	5040
$a(AU)$	0.031
$T_{int}(K)$	100
$\log(g)(\text{cm}^2 \text{s}^{-2})$	3.341
$\gamma_{v1}$	$1.58 \times 10^{-1}$
$\gamma_{v2}$	$1.58 \times 10^{-1}$
$\kappa_{IR}(\text{cm}^2 \text{s}^{-1})$	$3 \times 10^{-2}$
$\alpha$	0.5
$\beta$	1.0
$f_{H2}$ (ppm)	$8.5 \times 10^5$
$f_{He}$ (ppm)	$1.5 \times 10^5$
$f_{H2O}$ (ppm)	370
$f_{CH4}$ (ppm)	1
$f_{CO}$ (ppm)	31.6
$f_{CO2}$ (ppm)	0.2

TABLE 2

GAUSSIAN PRIOR PARAMETER VALUES AND WIDTHS. THE TRUE STATE IS THE SAME AS IN TABLE 1 BUT IN LOGARITHMIC UNITS FOR SOME OF THE PARAMETERS. THE MIXING RATIOS OF EACH GAS,  $f_k$ , ARE IN PPM. THE INFRARED OPACITY,  $\kappa_{IR}$ , HAS UNITS OF  $\text{cm}^2 \text{g}^{-1}$ .  $\gamma_{v1}$ ,  $\gamma_{v2}$ ,  $\alpha$ , AND  $\beta$  ARE ALL UNIT-LESS. NOTE THAT WE RETRIEVE THE LOG OF ALL VALUES EXCEPT  $\alpha$  AND  $\beta$ .

Parameter	True	Prior State ( $x_{a,i}$ )	Prior Width ( $\sqrt{(S_{a,ii})}$ )
$\log(\gamma_{v1})$	-0.8	-0.9	1
$\log(\gamma_{v2})$	-0.8	-0.7	1
$\log(\kappa_{IR})$	-1.52	-2.0	0.5
$\alpha$	0.5	0.5	0.05
$\beta$	1	1	0.25
$\log(f_{H2O})$	2.568	2	6
$\log(f_{CH4})$	0.0	1	6
$\log(f_{CO})$	2.663	2	6
$\log(f_{CO2})$	-0.70	1	6

TABLE 3

NUMERICAL SUMMARY OF THE RETRIEVAL RESULTS FOR SEVERAL PARAMETERS AS DERIVED FROM EACH RETRIEVAL TECHNIQUE AND OBSERVATIONAL SCENARIO. FOR EACH PARAMETER AND EACH OBSERVATIONAL SCENARIO WE SHOW THE TRUE VALUE, THE  $1\sigma$  (68% CONFIDENCE INTERVAL) MARGINALIZED PRIOR UNCERTAINTIES, AND THE  $1\sigma$  MARGINALIZED UNCERTAINTIES DERIVED FROM OPTIMAL ESTIMATION (OE), BOOTSTRAP MONTE CARLO (BMC), AND DIFFERENTIAL EVOLUTION MARKOV CHAIN MONTE CARLO (DEMC) AS WELL AS THE AVERAGING KERNEL ELEMENT FOR THAT PARAMETER (A). THE GAS ABUNDANCES,  $f_i$ , ARE GIVEN IN TERMS OF VOLUME MIXING RATIO. WE ALSO SHOW A REPRESENTATIVE TEMPERATURE (100 MBARS TEMPERATURE) AND THE C-TO-O RATIO. THIS TABLE IS LAYED OUT SO THAT FOR A GIVEN PARAMETER EASY COMPARISONS IN EITHER THE OBSERVATIONAL SCENARIO (LEFT-RIGHT) OR THE RETRIEVAL TECHNIQUES (TOP-BOTTOM) CAN BE MADE.

Paramter	Broadband	Multi-Instrument	Future Telescope
$f_{H2O}$	True: $3.70 \times 10^{-04}$	$3.70 \times 10^{-04}$	$3.70 \times 10^{-04}$
	Prior: $4.94 \times 10^{-10} - 3.92 \times 10^{-03}$	$4.94 \times 10^{-10} - 3.92 \times 10^{-03}$	$4.94 \times 10^{-10} - 3.92 \times 10^{-03}$
	OE: $1.25 \times 10^{-06} - 7.82 \times 10^{-03}$	$3.68 \times 10^{-06} - 7.03 \times 10^{-04}$	$2.31 \times 10^{-04} - 4.56 \times 10^{-04}$
	BMC: $5.74 \times 10^{-05} - 2.68 \times 10^{-04}$	$1.18 \times 10^{-05} - 2.45 \times 10^{-04}$	$2.44 \times 10^{-04} - 4.03 \times 10^{-04}$
	DEMC: $1.84 \times 10^{-07} - 1.24 \times 10^{-03}$	$7.06 \times 10^{-06} - 1.40 \times 10^{-03}$	$2.83 \times 10^{-04} - 4.98 \times 10^{-04}$
	A: 0.872	0.983	0.999
$f_{CH4}$	True: $1.00 \times 10^{-06}$	$1.00 \times 10^{-06}$	$1.00 \times 10^{-06}$
	Prior: $2.63 \times 10^{-10} - 2.55 \times 10^{-03}$	$2.63 \times 10^{-10} - 2.55 \times 10^{-03}$	$2.63 \times 10^{-10} - 2.55 \times 10^{-03}$
	OE: $8.89 \times 10^{-11} - 5.73 \times 10^{-04}$	$4.54 \times 10^{-09} - 3.92 \times 10^{-05}$	$7.50 \times 10^{-07} - 1.54 \times 10^{-06}$
	BMC: $6.40 \times 10^{-08} - 1.64 \times 10^{-06}$	$9.07 \times 10^{-08} - 3.89 \times 10^{-06}$	$7.62 \times 10^{-07} - 1.46 \times 10^{-06}$
	DEMC: $2.21 \times 10^{-11} - 9.27 \times 10^{-07}$	$4.40 \times 10^{-11} - 1.96 \times 10^{-06}$	$7.39 \times 10^{-07} - 1.52 \times 10^{-06}$
	A: 0.259	0.979	0.999
$f_{CO}$	True: $3.16 \times 10^{-05}$	$3.16 \times 10^{-05}$	$3.16 \times 10^{-05}$
	Prior: $4.94 \times 10^{-10} - 3.92 \times 10^{-03}$	$4.94 \times 10^{-10} - 3.92 \times 10^{-03}$	$4.94 \times 10^{-10} - 3.92 \times 10^{-03}$
	OE: $4.97 \times 10^{-10} - 2.43 \times 10^{-03}$	$7.40 \times 10^{-10} - 4.13 \times 10^{-03}$	$5.87 \times 10^{-06} - 3.13 \times 10^{-05}$
	BMC: $2.05 \times 10^{-06} - 6.10 \times 10^{-05}$	$3.00 \times 10^{-07} - 1.81 \times 10^{-05}$	$4.37 \times 10^{-06} - 2.76 \times 10^{-05}$
	DEMC: $2.00 \times 10^{-10} - 5.32 \times 10^{-04}$	$8.89 \times 10^{-11} - 7.16 \times 10^{-05}$	$3.52 \times 10^{-06} - 2.66 \times 10^{-05}$
	A: 0.316	0.176	0.996
$f_{CO2}$	True: $2.00 \times 10^{-07}$	$2.00 \times 10^{-07}$	$2.00 \times 10^{-07}$
	Prior: $2.63 \times 10^{-10} - 2.55 \times 10^{-03}$	$2.63 \times 10^{-10} - 2.55 \times 10^{-03}$	$2.63 \times 10^{-10} - 2.55 \times 10^{-03}$
	OE: $7.73 \times 10^{-11} - 1.78 \times 10^{-04}$	$5.07 \times 10^{-09} - 4.25 \times 10^{-03}$	$1.94 \times 10^{-07} - 4.82 \times 10^{-07}$
	BMC: $9.44 \times 10^{-09} - 3.61 \times 10^{-07}$	$1.64 \times 10^{-09} - 1.10 \times 10^{-07}$	$2.03 \times 10^{-07} - 4.29 \times 10^{-07}$
	DEMC: $2.21 \times 10^{-11} - 9.27 \times 10^{-07}$	$2.33 \times 10^{-11} - 7.37 \times 10^{-07}$	$2.30 \times 10^{-07} - 5.06 \times 10^{-07}$
	A: 0.508	0.689	0.999
$T_{100mb}[K]$	True: 1313	1313	1313
	Prior: 876 – 1503	876 – 1503	876 – 1503
	OE: 932 – 1358	1075 – 1274	1267 – 1340
	BMC: 1150 – 1249	1117 – 1284	1278 – 1327
	DEMC: 1050 – 1355	1136 – 1374	1294 – 1348
$C/O$	True: $8.00 \times 10^{-2}$	$8.00 \times 10^{-2}$	$8.00 \times 10^{-2}$
	Prior: $3.82 \times 10^{-02} - 8.00$	$3.82 \times 10^{-02} - 8.00$	$3.82 \times 10^{-02} - 8.00$
	OE: $2.07 \times 10^{-03} - 1.00$	$6.26 \times 10^{-03} - 0.938$	$2.25 \times 10^{-2} - 8.93 \times 10^{-2}$
	BMC: $1.60 \times 10^{-02} - 0.32$	$1.29 \times 10^{-02} - 0.427$	$1.81 \times 10^{-02} - 8.09 \times 10^{-2}$
	DEMC: $3.53 \times 10^{-04} - 0.97$	$1.93 \times 10^{-04} - 0.721$	$1.27 \times 10^{-02} - 6.90 \times 10^{-2}$

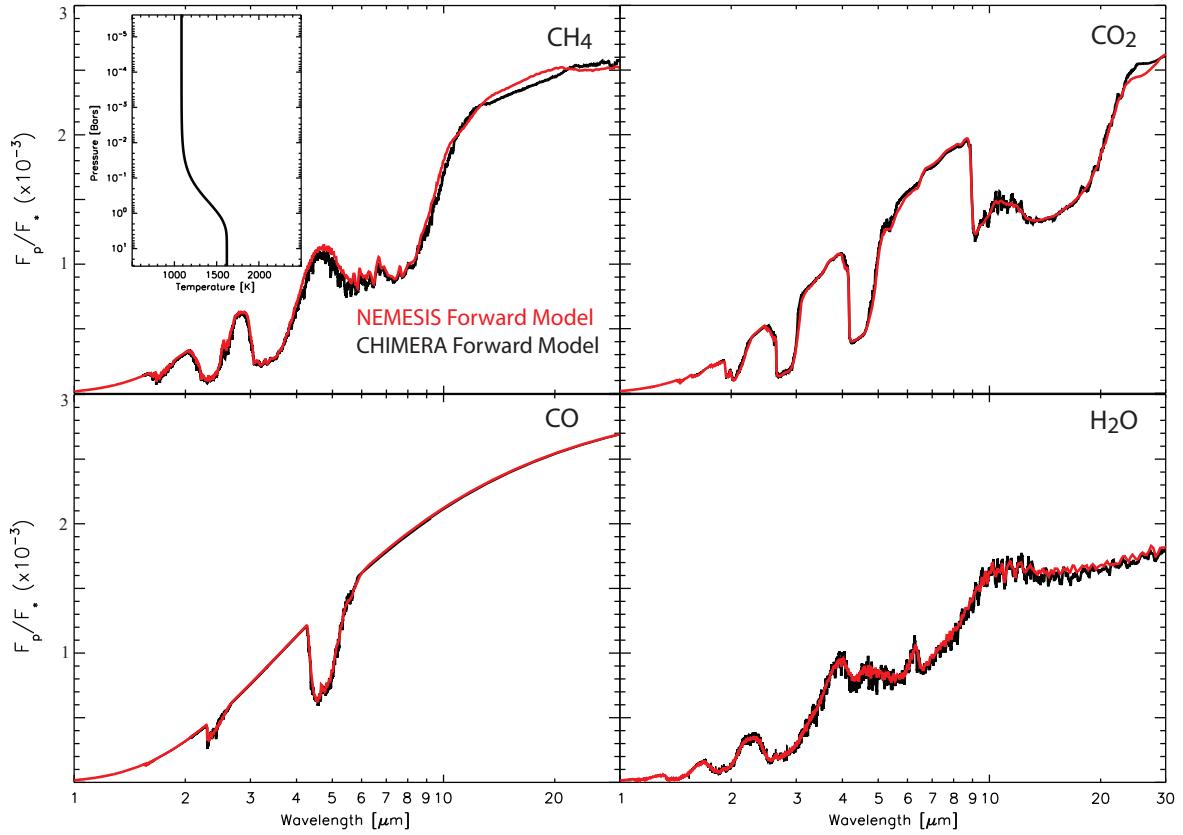


FIG. 1.— Comparison of the thermal emission spectrum from our forward model (black) with the NEMESIS forward model (red). The temperature-pressure profile is shown in the inset. For this comparison we assume uniform mixing ratios of  $10^{-4}$  for CH<sub>4</sub>, CO<sub>2</sub>, CO, and H<sub>2</sub>O. H<sub>2</sub> is set to 0.85 and He is set to 0.15. This planet is assumed to be hydrogen dominated (mean molecular weight of 2.3 amu) with a radius of 1R<sub>J</sub>, a gravity of 22 ms<sup>-2</sup> orbiting a 5700 K pure blackbody star with a radius of 1R<sub>sun</sub>.

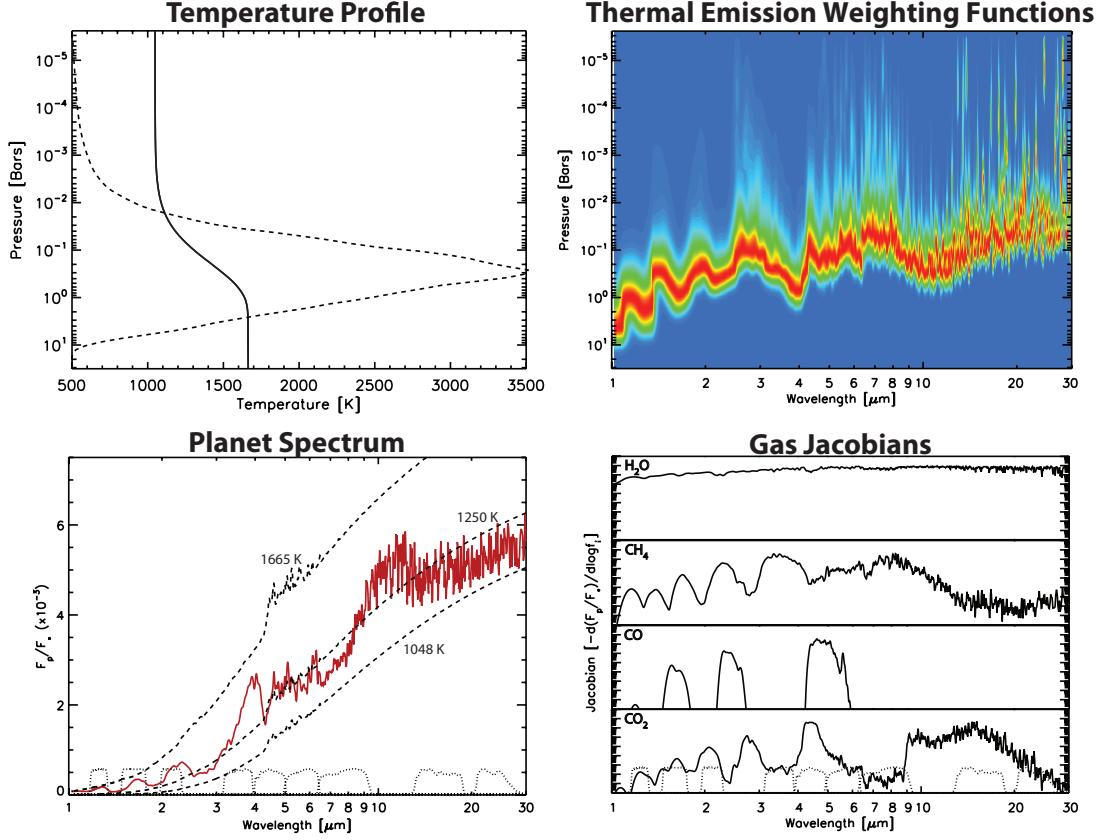


FIG. 2.— Synthetic planet atmosphere and spectrum. Top-Left: Model temperature-pressure profile. The solid curve is the temperature profile and the dashed curve is the averaged thermal emission contribution function, or where the emission in the atmosphere is coming from. The temperature profile is constructed using equations 13-16 and the parameters in Table 1. Top-Right: Thermal emission contribution function. This plot shows where the emission is coming from as a function of wavelength, smoothed to a resolution of 0.05 microns. Red corresponds to the peak of the thermal emission weighting functions, where the optical depth is unity, and blue represents zero emission. Most emission emanates between a few bars and 0.01 bars with deeper layers probed by shorter wavelengths. Bottom-Left: Resulting spectrum smoothed to a resolution of 0.05 microns. Blackbodies for the hottest, coolest, and average temperatures are shown. The dotted curves at the bottom are the filter profiles for typical photometric observations. Bottom-Right: Gas Jacobian generate from equation 9. This plot shows the sensitivity of the flux contrast as a function of wavelength to the various absorbers (the units are arbitrary but consistent).

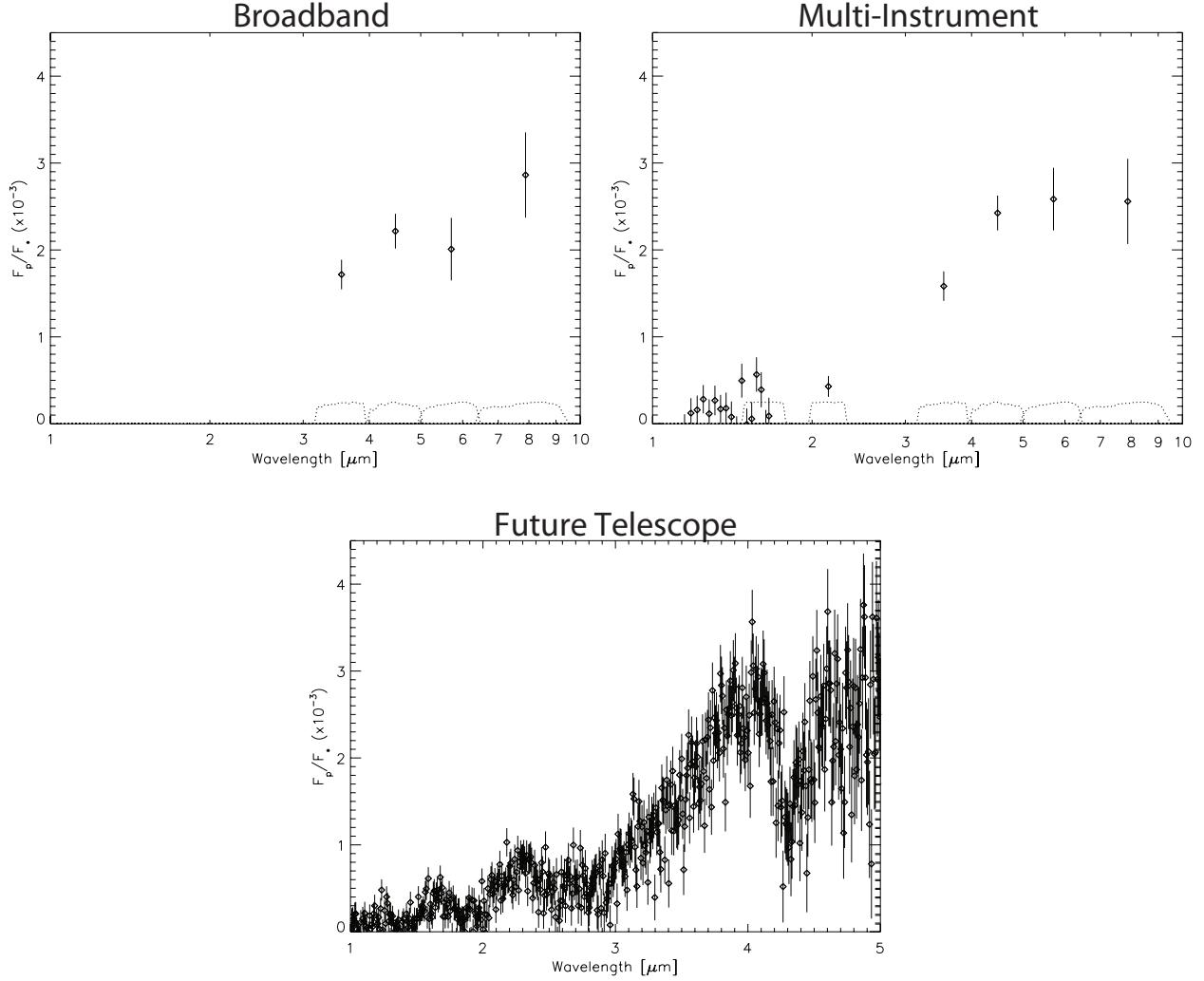


FIG. 3.— The spectrum of the synthetic hot jupiter observed in three different scenarios. These “observations” are created by convolving the high resolution spectrum in Figure 1 with the appropriate instrumental profiles. Random noise is then added to each data point. Top-Left: Synthetic observations as viewed through the Spitzer broadband 3.6, 4.5, 5.7, and 8  $\mu\text{m}$  channels. Top-Right: Multi instrument observations that include WFC3 (1.15-1.63  $\mu\text{m}$ ), ground based H and K<sub>s</sub>, and Spitzer Broadband (3.6, 4.5, 5.7, and 8  $\mu\text{m}$ ). Bottom: Hypothetical future spaceborne observations. The dotted curves on the bottom of each plot are the photometric transmission functions.

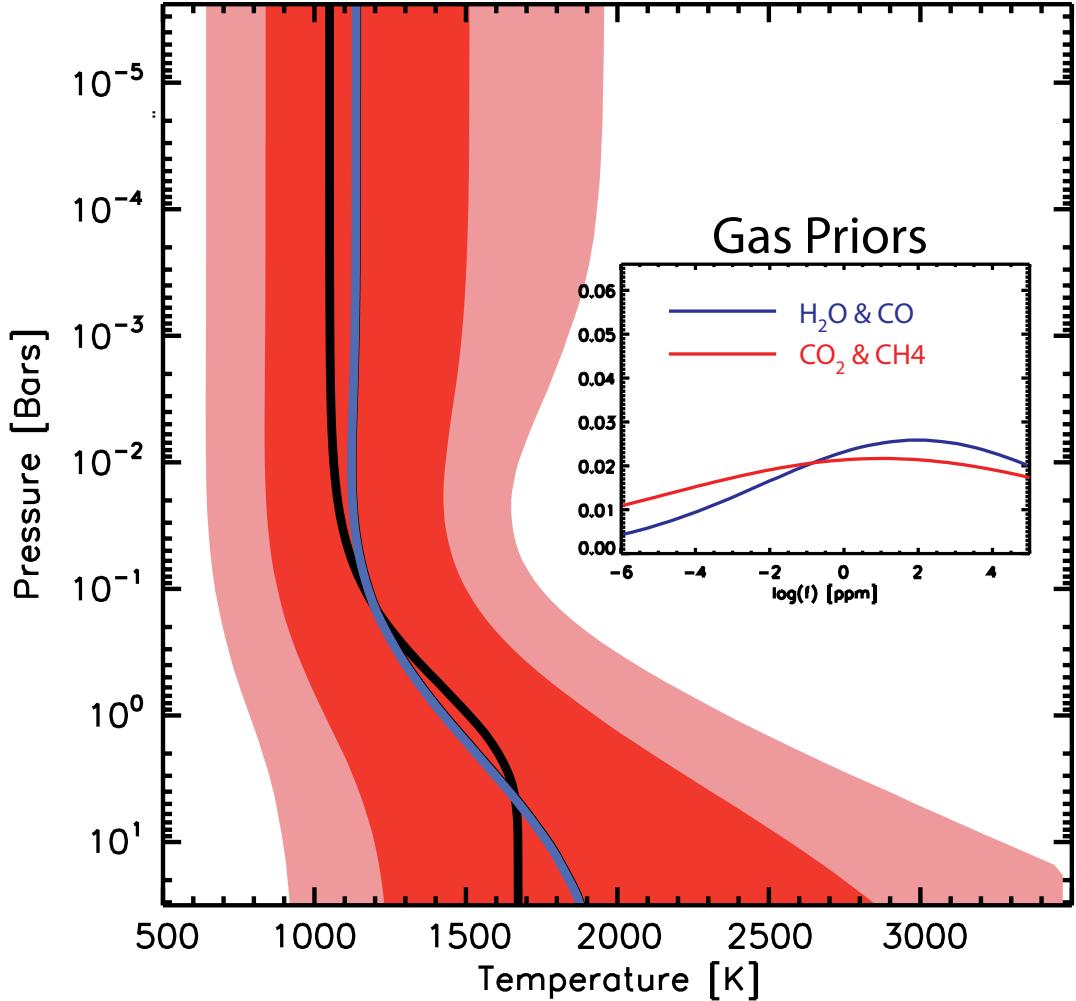


FIG. 4.— Temperature and gas priors (inset). Dark red represents the  $1\sigma$  spread in the allowed temperature profiles as a result of the prior parameter distributions in Table 2. Light red is the  $2\sigma$  spread allowed in the temperature profiles. The blue curve is the median temperature profile and the black curve is the temperature profile constructed from  $\mathbf{x}_a$  in Table 2. The gas priors are broad Gaussians.  $H_2O$  and  $CO$  have the same prior mean,  $CH_4$  and  $CO_2$  have the same prior mean. Note that the prior is gaussian in  $\log$  of the mixing ratios. The y-axis on the inset is normalized probability.

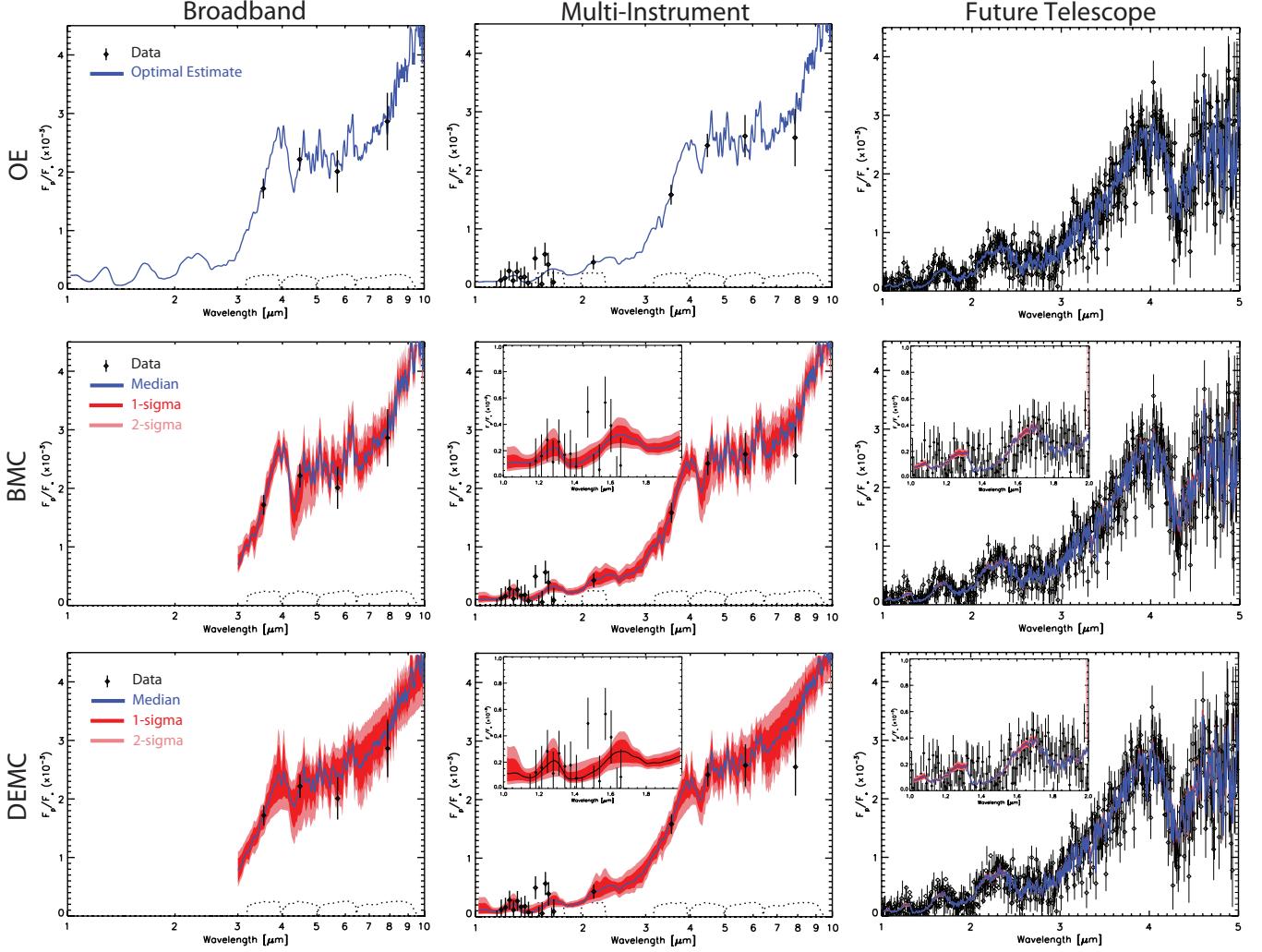


FIG. 5.— Fits to the three different sets of data (columns) from each of the three different retrieval techniques (rows). The first scenario consists of the four IRAC photometry channels. The second scenario consists of the four IRAC photometry channels, ground based H and K<sub>s</sub> band photometry, and HST WFC3 spectroscopy. The third scenario is representative of a FINESSE-like future, spaceborne telescope. For the Optimal Estimation, the single best fit is shown in blue. The Bootstrap Monte Carlo and the Differential Evolution Markov Chain Monte Carlo approaches generate many thousands of spectra. The median of these spectra is shown in blue and the 1- and 2 $\sigma$  spread in the spectra are shown in dark and light-red, respectively. The dotted curves at the bottom of each panel are the broadband filter transmission functions. The insets are a zoom in of a spectral region between 1 and 2  $\mu\text{m}$  to better show the spread in the spectra. Note that there is virtually no spread in the spectra for the future telescope case.

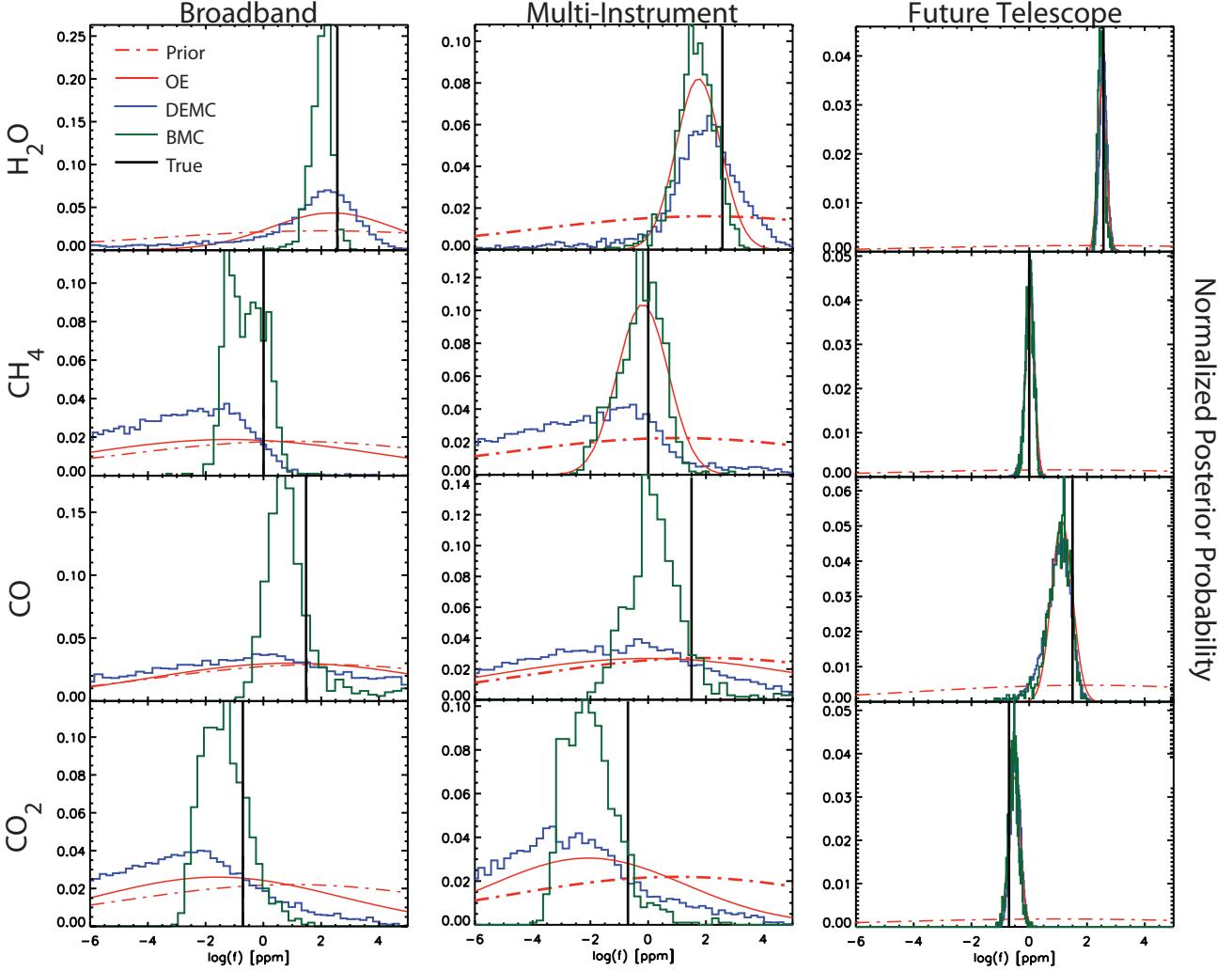


FIG. 6.— Marginalized posterior probability distributions for each of the retrieved gases (rows) and observational scenario (columns). In each panel the probability distribution for each retrieval technique are shown in different colors. The Gaussian probability distributions from Optimal Estimation are in red, Differential Evolution Markov Chain Monte Carlo in blue, and Bootstrap Monte Carlo in green. The priors for each gas are the dot dashed red curve. The true answer is the vertical black line.

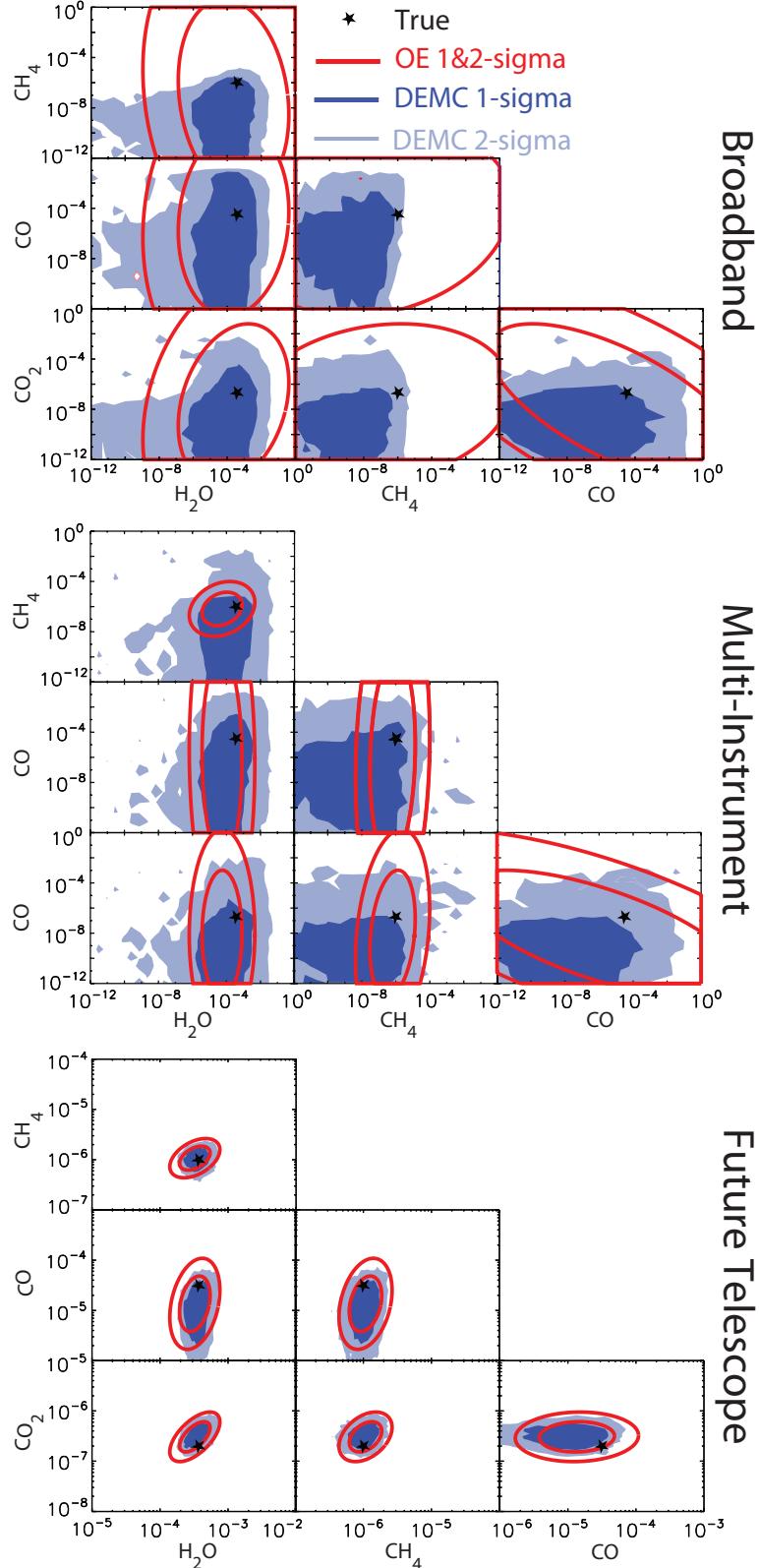


FIG. 7.— Gas correlations for each of the observing scenarios. The red curves in each are the analytic confidence intervals from the Optimal Estimation posterior covariance matrix ( $\hat{\mathbf{S}}$ ). The inner ellipses are the  $1\sigma$  (68%) and the outer ellipses are the  $2\sigma$  (95%) confidence interval. The  $1\sigma$ - and  $2\sigma$  confidence intervals derived from the Differential Evolution Markov Chain Monte Carlo are shown in dark and light blue, respectively. Note that the scales for the confidence intervals derived from the broadband observations (top) and the multi-instrument observations (middle) are the same. The scale on the future telescope (bottom) confidence intervals is much smaller.

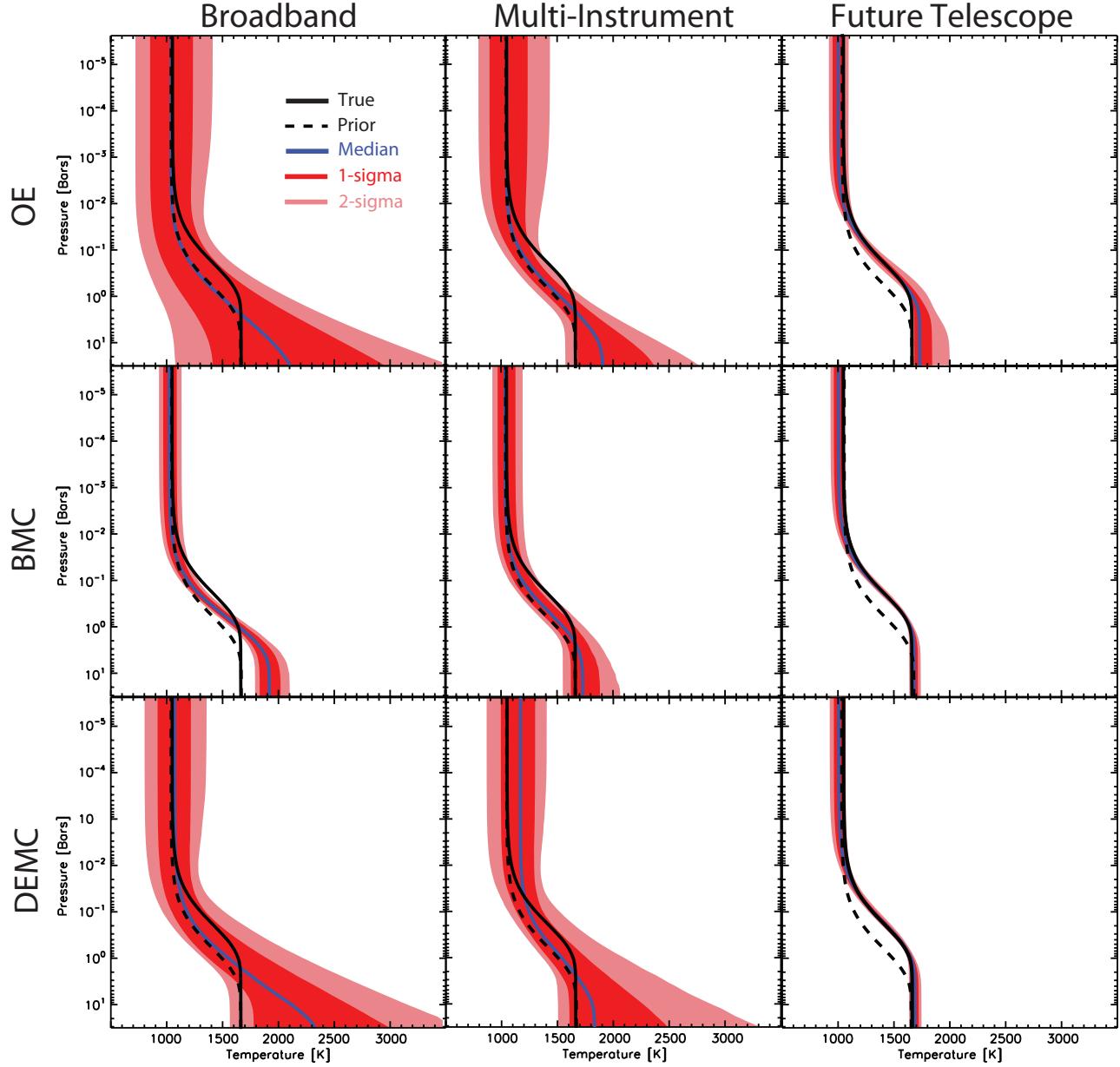


FIG. 8.— Temperature profile posteriors for each observational scenario (columns) and each retrieval technique (rows). The solid black curve in each panel is the true temperature profile constructed with equations 13–16 and the parameters in Table 1. The dashed black curve is constructed from the temperature parameters,  $x_a$ , just as in Figure 4. The blue curve is the median temperature profile. The dark and light red regions are the 1- and 2 $\sigma$  (68% and 95%) uncertainties, respectively.

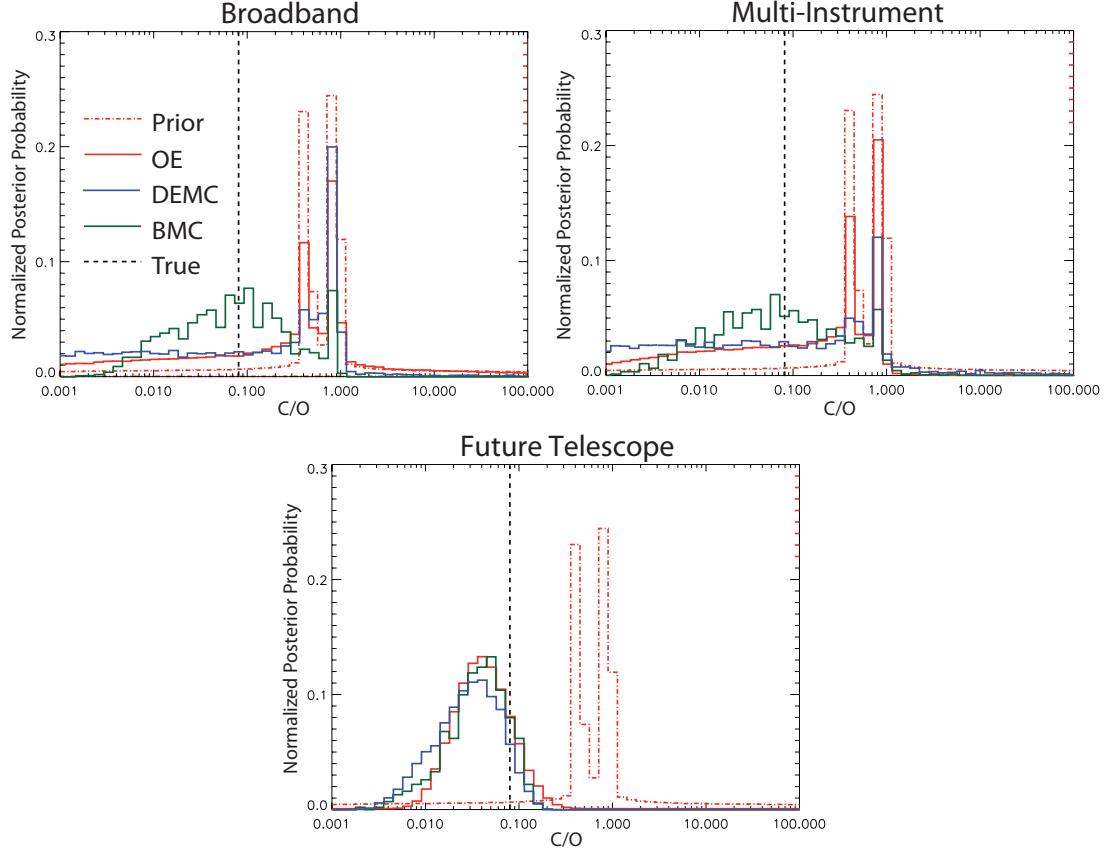


FIG. 9.— C to O ratio posteriors. The dot-dashed green curve is the prior, the solid red curve is from OE, blue is from DEMC, and green is from BMC. The vertical dashed line is the true C/O. The top left panel are the C/O's derived from the broadband observational scenario, top right are the C/O's derived from the multi-instrument scenario, and the bottom are the C/O's derived from the future spaceborne telescope scenario. Though it appears that the BMC characterizes the C/O errors well, it is for the wrong reasons. See §3.3.3

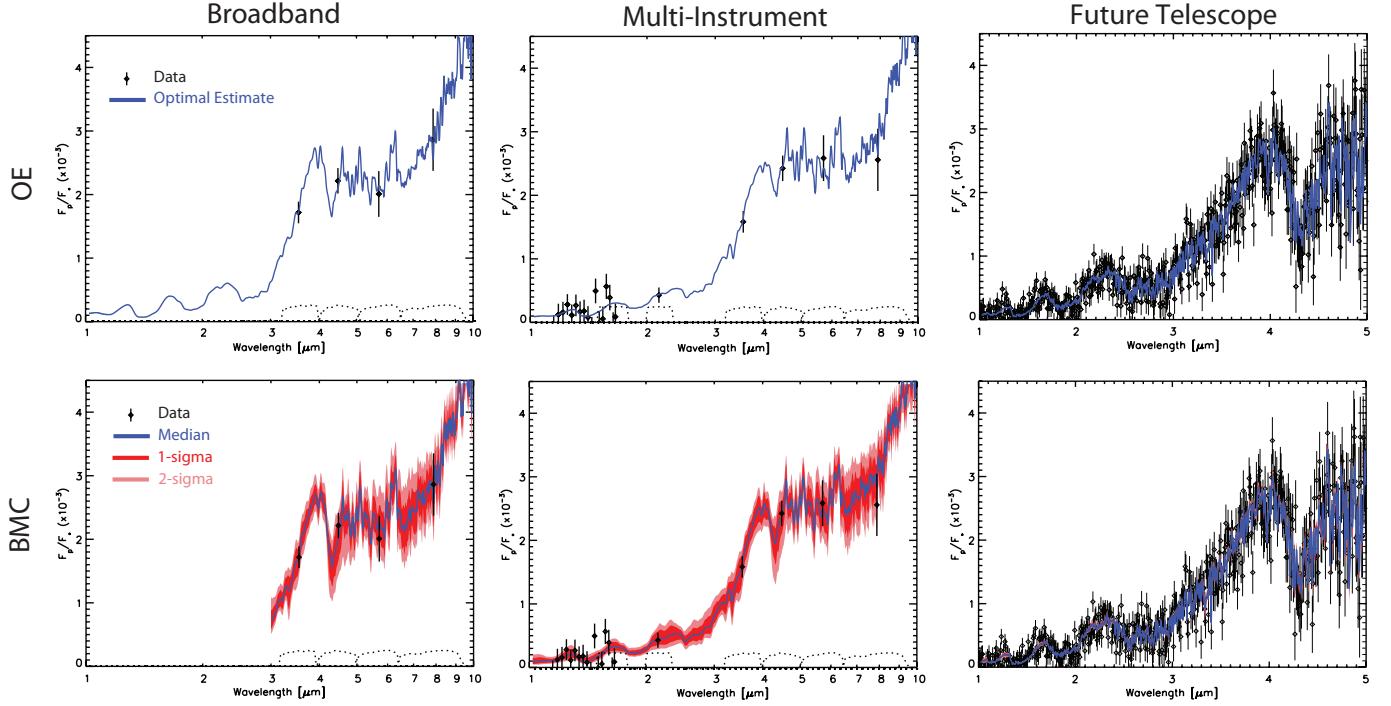


FIG. 10.— Fits to the three different sets of data (columns) from two of the retrieval techniques (rows) using the Level-by-Level temperature approach. For the optimal estimation approach, the single best fit is shown in blue. The Bootstrap Monte Carlo approach generates many thousands of spectra. The median of these spectra is shown in blue and the 1- and  $2\sigma$  spread in the spectra are shown in dark and light-red, respectively. The dotted curves at the bottom of each panel are the broadband filter transmission functions.

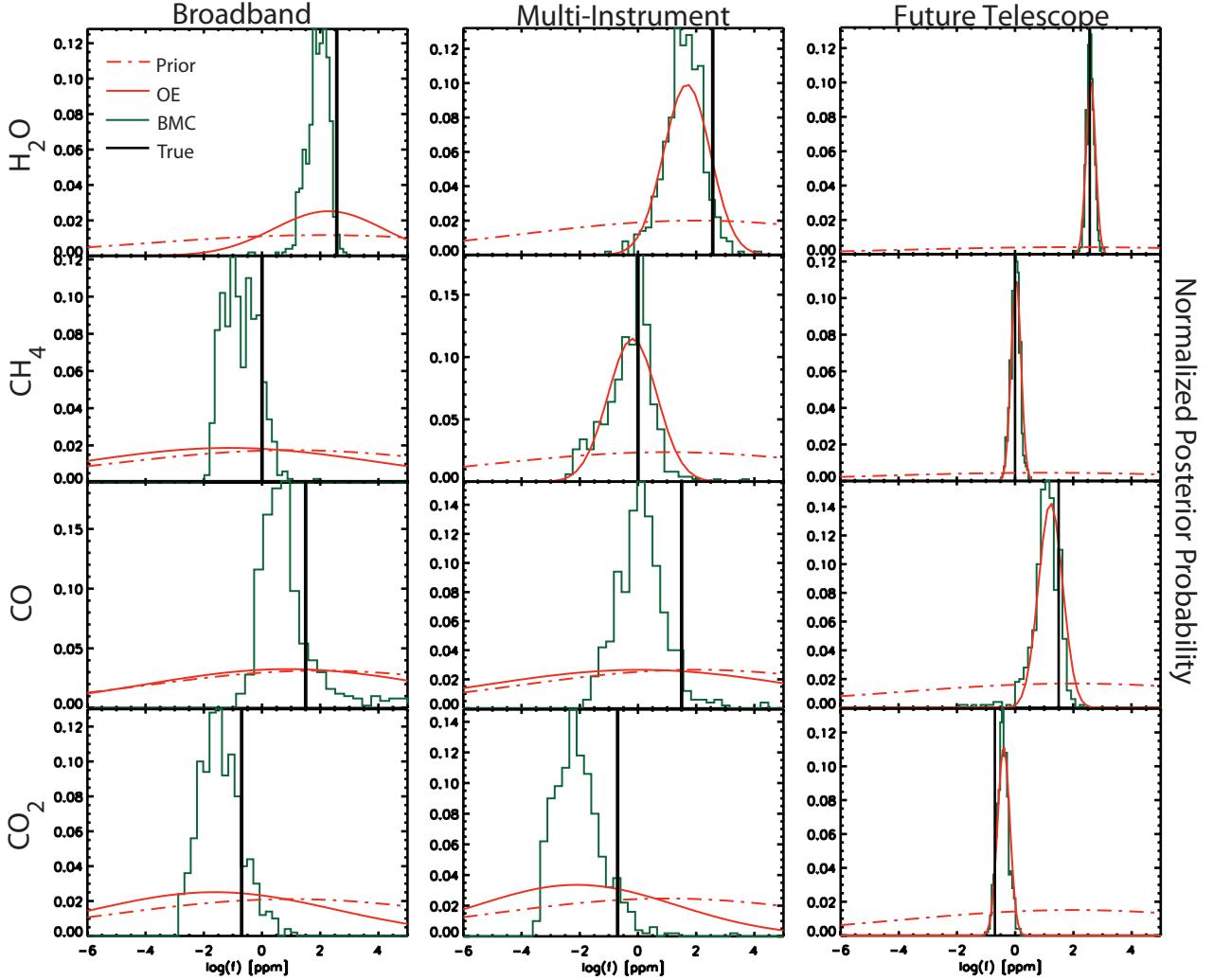


FIG. 11.— Marginalized posterior probability distributions for each of the retrieved gases (rows) and observational scenario (columns) using the Level-by-Level temperature profile approach. In each panel the posteriors for optimal estimation (red) and Bootstrap Monte Carlo (green) are shown. The Gaussian priors for each gas are shown with the dot dashed red curve. The true answer is the vertical black line.

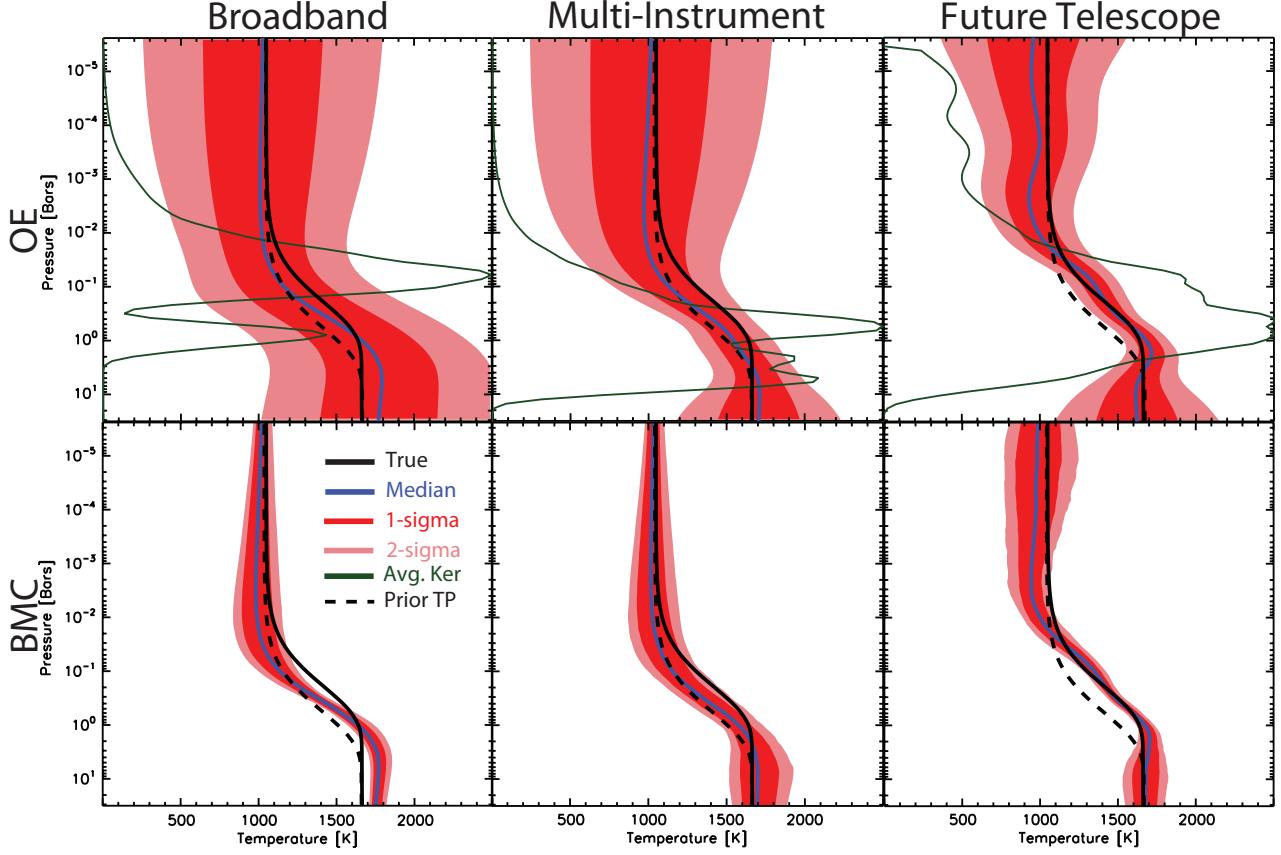


FIG. 12.— Temperature profile posteriors using the Level-by-Level temperature profile approach for each observational scenario (columns) and two of the retrieval technique (rows). The solid black curve in each panel is the true temperature profile constructed with equations 13-16 and the parameters in Table 1. The blue curve is the median temperature profile. The dashed black curve is the prior temperature profile constructed from  $\mathbf{x}_a$ , as in Figure 4. The prior widths for each level (not shown) are  $\pm 400$  K. The dark and light red regions are the 1- and 2 $\sigma$  (68% and 95%) uncertainties, respectively. The green curve is the averaging kernel profile for temperature. The atmospheric regions over which this is a maximum is where we can retrieve temperature information with less dependence on the prior (see text).

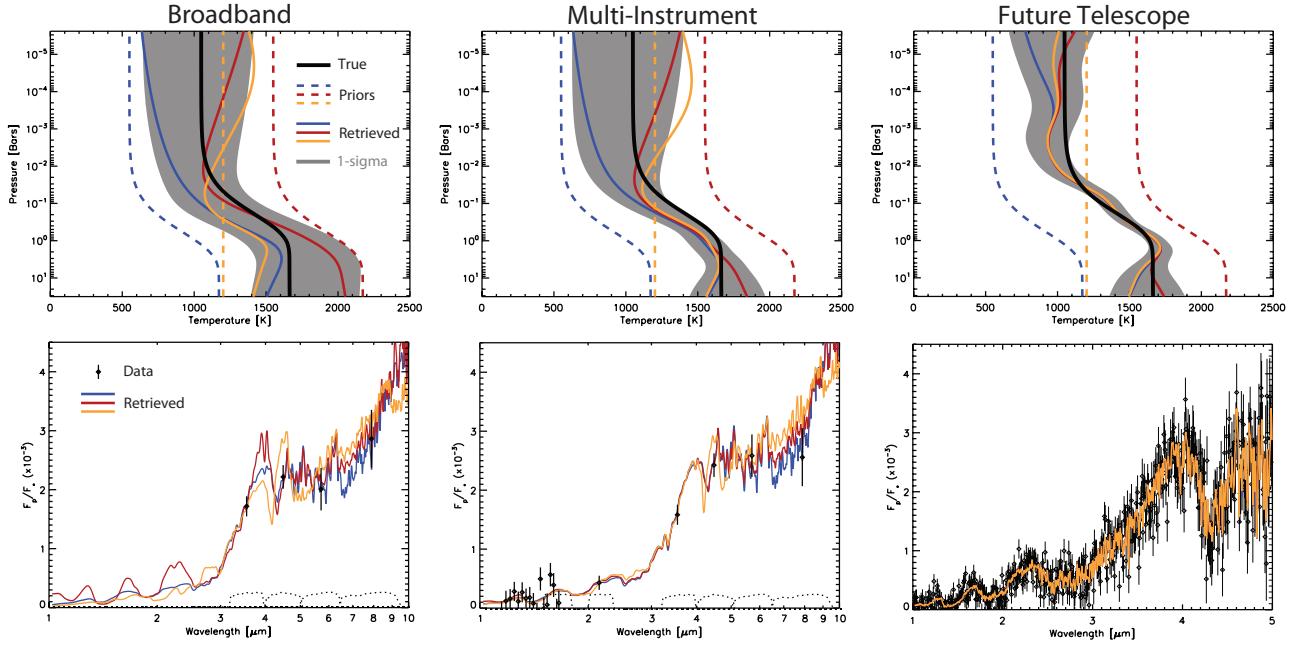


FIG. 13.— The effect of three different Level-by-Level temperature profile priors on the retrieved temperatures (top) and spectra (bottom). The different temperature profile priors are shown as the colored dashed curves. The prior widths (not shown) at each level are  $\pm 400$  K, similar to those in Figure 12. The resultant retrieved profiles are shown as the solid colored curves. The thick black curve is the true temperature profile. The solid grey region is the  $1\sigma$  confidence interval from the retrievals in Figure 12. Note how the retrieved profiles all converge within the  $1\sigma$  confidence interval regardless of the temperature prior. The best agreement is in the middle atmosphere where the thermal emission weighting functions are a maximum, and hence the averaging kernel profiles from Figure 12 are also a maximum. The spectra in the second row illustrate the effects of the different retrieved temperature profiles of corresponding color. There is virtually no difference in the resultant spectra for high quality data. The dotted curves at the bottom are the broadband filter functions.